Second Series *April 15, 1932* *Vol. 40, No. 2*

THE

PHYSICAL REVIEW

X-RAY STUDIES ON Bi SINGLE CRYSTALS

By A. Goetz and R. C. Hergenrother

Norman Bridge Laboratory of Physics, California Institute of Technology

(Received February 2, 1932)

Abstract

An x-ray analysis of different types of Bi single crystals is described in order to decide whether or not the difference in density, electric conductivity and thermal e.m.f. observed between Bi crystals grown within (M), or outside (N) of a transverse magnetic field were caused by a change in lattice spacing or due to an alteration of the lattice parameter.

Determination of the interplanar spacing along (11$\bar{1}$). The Bi crystals were grown by the method described by Goetz, and the cleaved planes (111) and (11$\bar{1}$) were used. By means of the precision method of Siegbahn the interplanar spacing along (111) was determined with Mo $K\alpha$ radiation as 3.9453 ± 0.0002A at 23°C. No difference larger than 0.003 percent of $d_{(111)}$ was found to exist between N and M crystals hence the former conjecture was proved that the changes caused within the crystal by its formation in a magnetic field affect only its mosaic structure and do not influence the lattice.

Determination of the parameter. Two different methods have been used to measure the integrated reflected intensities for different orders. By means of a new photographic recording method the intensities of the first to the fifth order have been compared and no consistent change of the parameter could be found. In addition to the photographic records an ionization method of high sensitivity has been used involving a new type of x-ray spectrometer—designed by Hergenrother—in connection with a Hoffmann electrometer. The photographic results were verified and the integrated intensities could be measured up to the seventh order, thus permitting much greater accuracy in the determination of the parameter than has been previously reported. Making use of Pauling and Sherman's computation of the atomic structure factor and applying an experimentally determined value of the Debye factor, the parameter along (111) was found to be $0.359\text{A} \pm 0.0006\text{A}$, which value lies well within the limits of Hassel and Mark and is slightly larger than James' observation.

Conclusion. The non-existence of a change in lattice structure and in parameter for the N and M crystals shows that the electric volume effects which distinguish the two types of crystals can exist in spite of the unchanged conditions for the existence and distribution of the "free" electrons as given by the geometric configuration of the atoms in the lattice. Hence the cause of these effects can only be ascribed to changes of certain variations in the lattice, the period of which must be many times larger than the size of an elementary rhomboheder in order to be undetectable to x-ray analysis. This implies the assumption of very large "mean free paths" for the electrons in a crystal.

Introduction

IN PREVIOUS papers certain changes have been described by Goetz and Hasler[1] and Goetz and Focke[2] which a bismuth single crystal undergoes, if it is crystallized in a transversel magnetic field. It has been observed that for an orientation in which the principal axis (111) was parallel to the rod and normal to the lines of force a large thermal e.m.f. existed between a section of the crystal grown without and another crystallized within the field, although the two sections did not show a change in orientation larger than 0.5°. The thermal e.m.f. changed sign and magnitude in the opposite orientation, where the (111) axis was normal to the axis of the rod and normal to the lines of force. In addition to this effect a change in conductivity and density was observed, such that in the former case the density decreased (*ca.* 0.3 percent below normal), whereas it was found to increase in the latter orientation (*ca.* 0.2 percent above normal).

Obviously the question arose as to whether or not the genesis of the crystal under the influence of a magnetic field caused a change in the interplanar spacing of the crystal lattice or in its parameter, of which possibilities the former seemed to be indicated by the change in density.

Preparation and Origin of the Crystals

The bismuth crystals used were grown one half within and the other half outside of a magnetic field of approximately 13,000 Gauss using the method previously described.[3] The orientation of the crystals was either: (111) parallel or normal to the axis of the rod, the diameter was 3–4 mm, the length 10–12 cm. The metal used was of the origin "B"[4] and showed the largest magnetic effect.

Since it is rather difficult to prepare good cleavage surfaces suitable for intensity measurements—as is necessary for the determination of the parameter—a number of experimental difficulties had to be overcome,* and a large number of cleavages had to be performed in each case before a satisfactory plane could be obtained.

[1] A. Goetz and M. F. Hasler, Phys. Rev. **36,** 1752 (1930).

[2] A. Goetz and A. B. Focke, Phys. Rev. **37,** 1044 (1930).

[3] A. Goetz, Phys. Rev. **35,** 193 (1930).

[4] Reference 1, page 1775, Table IV.

The "B" metal (origin: Hartmann and Braun, Frankfurt a.M. Germany) contains according to a spectroscopic analysis by Mr. M. F. Hasler, approximately the following impurities:

0.01% Cu
0.028% Ag
0.061% Pb
0.0022% Tl
0.0013% Te.

* The best technique seems to be to cleave at room temperature (liquid air was tried, however, without better success) with a very thin Gillette blade. For this purpose the crystal is held with wax to a piece of hard wood and the blade adjusted to the cleavage plane is struck sharply with a small hammer. Two pieces of rubber as thick as the crystal support the ends of the blade in order to prevent it from cutting deep into the crystal. Hence the blade penetrates into the crystal only to approximately one fifth of its diameter, whereas the rest of the cross-

Determination of D Along [111]

After preliminary measurements with the Bragg method had proved the necessity of high accuracy, the final observations were taken with a Siegbahn spectrometer[5] at which the angle scale could be read to 30 inches with a pair of vernier scales and microscopes. A water-cooled x-ray tube of Müller type with molybdenum target was used as source and was run at 57 kv and 10 m.a. with half-wave rectification (kenetron).

In order to be able to compare the reflections from "magnetic" and normal crystals a special holder shown in Fig. 1 was made. Two cleaved sections of the same crystal, C_1 and C_2,—one of the magnetic and one of the normal half—were cemented onto the holders a and b. The crystals were adjusted parallel to each other by putting them on a plane glass plate with the faces

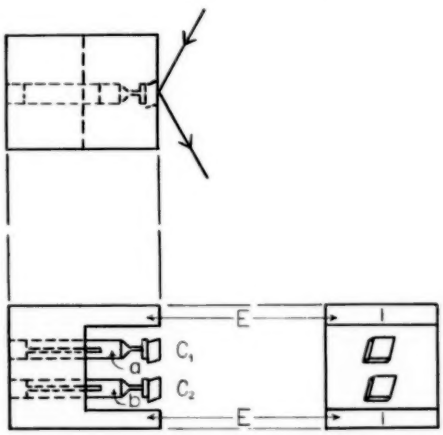

Fig. 1. Design of the double crystal holder for the photographic method in order to compare M and N crystals simultaneously.

downward, then the holder was put over the crystals standing with its edges EE on the plate. After this the sliding pins a and b were lowered until they gently touched the crystals. Then the crystals were attached to the tips of a and b with a drop of a non-shrinking cement. The holder itself could be mounted onto a movable stage in order to align the crystal faces with the axis of rotation of the spectrometer. The final adjustment was checked by optical reflections in the usual way.

Using a precision method developed by one of us for this purpose (6) on the base of Siegbahn's technique, the reflections were recorded on a photographic plate.

Two reflections of a given order were obtained for each crystal, according to two opposite corresponding positions of the crystal including the known

section splits without having been touched by the blade. The region which came into touch with the blade is distorted and twinned whereas the large portion of the surface is very flat. This procedure can be applied to the planes (111) and ($11\bar{1}$), though the latter are very difficult to obtain in good perfection.

[5] Siegbahn, Spektroskopie der Röntgenstrahlen p. 64, Berlin 1924.

angle ϕ. The separation of each set of lines on the plate $\delta = R \cdot d\phi'$, is determined by the comparator and by measuring the distance R between the axis of the spectrometer and the plate. The glancing angle θ is then determined by:

$$\theta = \tfrac{1}{4}\left(\phi + \frac{360\delta}{2\pi R}\right) \cdots .$$

Fig. 2 shows a reproduction of the 5th order reflection of Mo $K\alpha$ doublet reflected from [111] of a "normal" (N) and a "magnetic" (M) crystal. The separation of the $K\alpha_1$-lines, measured as an average from several plates was 0.799 mm for "normal" and 0.777 mm for "magnetic" crystals. Since this difference was within the uncertainty of the measurement of the center of each line the value 0.778 was chosen as an average. This proves that the interplanar spacing along [111] is *not* different between "magnetic" and "normal" sections of the crystals within the limits of error (0.003 percent).[6]

Fig. 2. Reproduction of the 5th order reflection of the Mo $K\alpha$-doublet from (111) of a "normal" and a "magnetic" crystal.

The values measured are:

$$\delta = 0.788 \pm 0.001 \text{ mm}$$
$$R = 172.6 \pm 0.2 \text{ mm}$$
$$\phi = 106.329° \pm 0.004°$$
$$\theta = 26.647_0 \pm 0.001°.$$

Hence the interplanar spacing along [111] is given by:

$$d = \frac{n \cdot \lambda}{2 \sin \theta} = 3.9453\text{A} \pm (0.0002\text{A}) \text{ at } 23°\text{C}.$$

This value is in good agreement with the results obtained by James[7] (3.95A $\pm$ 0.08A) and Hassel and Mark[8] (3.945A).

The possibility of a lattice deformation in another direction to the principal axis was checked also although the oriented thermal e.m.f. of "magnetic" crystals made such an effect less probable.[1] Reflections from (11$\bar{1}$) planes measured in the same manner gave also no change between two different sections. For these measurements the accuracy obtained was, however, only 0.5 percent due to the imperfect cleavage of such planes.

[6] R. C. Hergenrother, Physics **2,** 63 (1932).
[7] R. W. James, Phil. Mag. **42,** 193 (1921).
[8] O. Hassel and H. Mark, Zeits. f. Physik **23,** 269 (1924).

The conjecture already presented in a previous paper[1] that the change in density caused by the crystallization in the magnetic field be due to a change in the mosaic structure of the crystals seems therefore justified.

Determination of the Parameter

The identity of the lattices of the two types of Bi is however not established as long as the parameters are not compared, although a change in this magnitude would not account for the observed change in densities. A determination of the parameter involves an exact method to determine the reflected integrated intensities of different orders, and a photographic method as well as the direct measurement by means of ionization have been applied.*

The photographic method

The first attempt to determine the reflected integrated intensities by photometry of photographic records for different orders (1 to 5) showed a distinct difference between "magnetic" and "normal" portions of the crystal; the results however were too inconsistent to be relied upon in spite of a qualitative agreement with James' results. Table I presents the intensities in arbitrary units standardized for the second order (=100), for the first five orders of the simultaneous measurement of a "magnetic" and a "normal" section of the same crystal from three different plates in comparison with James'[7] values. Since this author used Rh $K\alpha$-radiation an exact agreement is not to be expected.

Table I.

Order	James'[1] $\lambda=0.612$A	Plate *A*		Plate *B*		Plate *C*		Average *A,B,C.*	
		N	*M*	*N*	*M*	*N*	*M*	*N*	*M*
1	29	47	17.6	32.7	26.7	—	—	39.8	22.2
2	100	100	100	100	100	100	100	100	100
3	24	20.5	37.6	25.8	37.7	27.3	28.6	24.5	38.2
4	Not given	0.0	2.1	1.3	1.5	0.0	3.0	1.3	2.2
5	Not given	0.0	6.5	0.0	5.1	0.0	4.5	0.0	5.36

With the wide range of reflected intensities for different orders the disadvantage of the usual photographic recording method is obvious and therefore a new technique was developed. Although the quality of the cleaved surfaces was not sufficient to permit a direct measurement of integrated intensities for the final determination of the parameter of Bi, the new method, which uses the photographic plate only as zero-instrument, shows certain general advantages, that may justify its description.

* The source of the x-radiation as well as the high tension arrangement used was essentially the same as described above, with the only difference that the filament of the x-ray tube was heated by a storage battery of large capacity, the constancy of the current being regulated manually by a micrometer rheostat of special design and being checked by the emission current accurate to 0.02 percent. Because of the large range of reflected x-ray intensities for different orders, two standard emission currents were use: 1 m.a. for low (first to third) and 10 m.a. for high orders (fourth to seventh). This made it possible to cover all ranges of intensity without necessitating a change in sensitivity of the electrometer for the ionization measurement nor requiring extremely short or long exposures for the photographic method. It involved however a correction for the leakage currents, since the batteries were on the high tension side.

The Siegbahn spectrometer as mentioned above was used for this purpose, though any spectrometer arranged for adjusting crystal as well as plate accurately in a predetermined position can be used. The method involves the following procedure. (Fig. 3.)

(a) In order to record a given order of Bragg reflection, the plate P is fixed at a predetermined position and the crystal C is turned at a definite angular speed through the Bragg-angle from a position $\theta-\phi$ to $\theta+\phi$ by

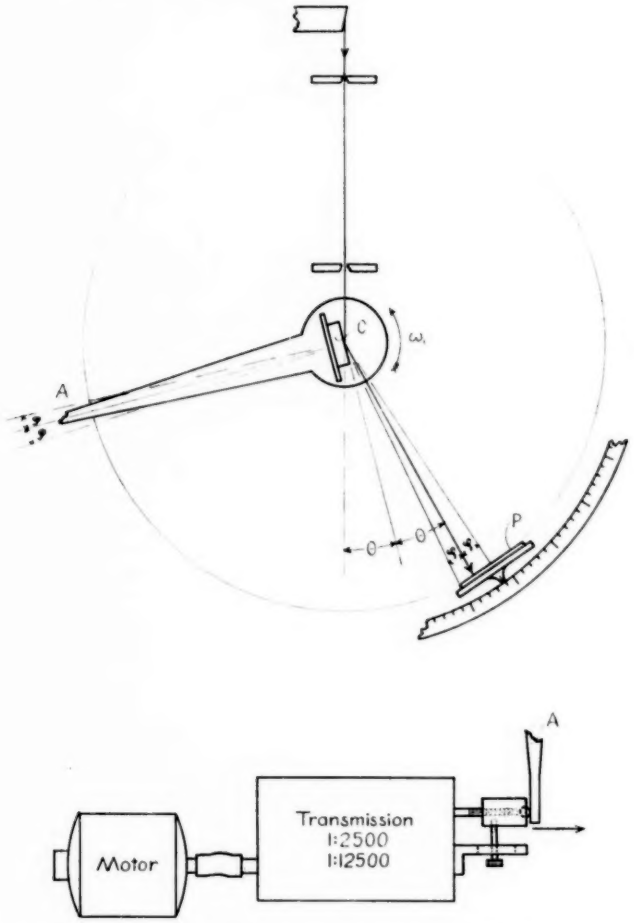

Fig. 3. Scheme of photographic recording method of intensities. θ = Bragg angle; $2\theta >$ rocking angle.

means of a motor driving a threaded shaft over a transmission, thus pushing the arm A of the crystal holder through the range 2ϕ. The motor has to be adjustable over a large range of speeds. The angle 2ϕ is slightly larger than the experimentally determined rocking angle of the crystal.

(b) To record successive orders of reflection, the plate is set at a series of successive positions which are chosen so that the respective reflected images are adjacent on the photographic plate.

(c) For different orders, the x-ray tube voltage and current are maintained the same, but the angular speed of turning of the crystal is set for each case so that the blackness of the recorded lines of the different orders are nearly the same on the photographic plate. (It is assumed that the relative intensities are known roughly from a previous measurement.) These angular speeds are so adjusted that the blackness of the line corresponds to the linear part of the plate characteristic.

(d) The plate is calibrated by turning the crystal to the order of reflection which is most intense and adjusting the crystal to a fixed position, where it gives a strong reflection. A number of exposures of this line are made for a series of different time intervals at successive positions of the plate.

The advantages over the usual photographic methods are the following ones:

(a) The total angular turning distance of the crystal is very small compared with the total Bragg angle (usual method) thus giving a great economy of time of exposure.

Fig. 4. Reproduction of the $K\alpha$-doublet reflections of the first five orders for N and M crystals on the (111) plane with monochromatic calibration lines.

(b) The lines of successive orders are close together, which insures uniform characteristics of the plate and uniform treatment in development and fixing of the photographic plate. Having the lines close together also allows quicker and more accurate microphotometer measurements.

(c) The adjustment of the angular speed of turning of the crystal can be done accurately by adjusting the speed of the driving motor. The relative opacities of the lines on the photographic plate can be determined with greater accuracy if the opacities are nearly alike and in the linear part of the plate characteristic than if they are widely different.

(d) The calibration of the plate is made with the same wave-length as the lines to be measured. The width and intensity distribution of the calibration lines is the same as that of the lines to be measured, which makes the measurement independent of response characteristics of the microphotometer.

Fig. 4 is a reproduction of a plate obtained with this method for the two types of crystals from (111). The different orders can easily be recognized by the increasing resolution by the $K\alpha$-doublet. In addition to the two sets of the first five orders there are the calibration lines. The results obtained from photometric measurements taking into account the speed ratios for the different orders and standardizing again for the $2d$ order for 4 different plates obtained from three different crystals are shown in Table II.

TABLE II.

Order	Plate D		Plate E		Plate F		Plate G	
	N	M	N	M	N	M	N	M
1	39.3	37.6	40.6	40.6	34.3	51.8	69.0	47.0
2	100	100	100	100	100	100	100	100
3	41.2	43.0	48.6	41.0	28.7	27.2	26.1	41.1
4	2.05	2.13	—	—	1.99	1.88	2.9	4.3
5	6.80	7.64	9.6	7.8	4.82	9.55	5.6	7.3

Although the data obtained are not sufficient for an exact determination of the absolute value of the parameter, the absence of a systematic difference supports the conclusion of its invariance with the two types of crystals.

The ionization method

In order to verify this and to determine the value of the parameter quantitatively with a higher degree of accuracy as had been reached before in previous investigations, the photographic method was abandoned and a special construction of an x-ray spectrometer was designed in order to record the reflected intensities by means of an ionization chamber and a Hoffmann electrometer as described separately by Hergenrother.[9]

With this method only one crystal at a time could be used which was either of the "normal" or the "magnetic" type and special great care had to be taken in order to obtain good cleavages.*

The intensities of the reflected $K\alpha$ doublet as a function of θ were measured repeatedly with different sets of crystals with great care, the values agreeing within *ca.* 1 percent for each set. Table III gives the data of the areas of the intensity curves together with the standardized values.

TABLE III.

Order	Normal		Magnetic	
	Area	Standardized	Area	Standardized
1	780	66	487	40.5
2	1180	100	1201	100
3	400	33.9	417	34.7
4	36.8	3.11	45.1	3.75
5	65.0	5.50	90.1	7.50

Comparison with the values of Table II show a still smaller variation of the intensities of N and M crystals, although the shape of the intensity curves showed evidently the presence of few facets, the existence of which affects the integrated intensities. The large difference in the first order is obviously

[9] R. C. Hergenrother, Physics **2**, April (1932).

* Since it is very difficult to obtain sufficiently flat surfaces for this kind of measurement it was tried to use natural surfaces of (111) by solidifying liquid Bi in a test tube until the first crystals appeared, then separating the liquid from the solid. The crystal-filigrane thus obtained was examined for individual crystals showing perfect (111) planes—a rare occurrence since the preferred growing plane of Bi is (100). Few perfect specimens could be found, but the operation of separating them from the neighboring crystals caused sufficient distortions to spoil them for intensity observations and all final measurements were therefore taken from cleaved crystals.

caused by surface distortion, being very influential due to the exceedingly small penetrating depth of the radiation for small angles. If only the respective maximum intensities are taken, their standardized values are found to agree within 2 percent.

It seems to be impossible to account for the remaining variations of the intensities by a change in parameter, since the variation is not systematic—neither in direction nor in size. Hence the conclusion was drawn definitely that *the parameter for "magnetic" and "normal" crystals is the same.*

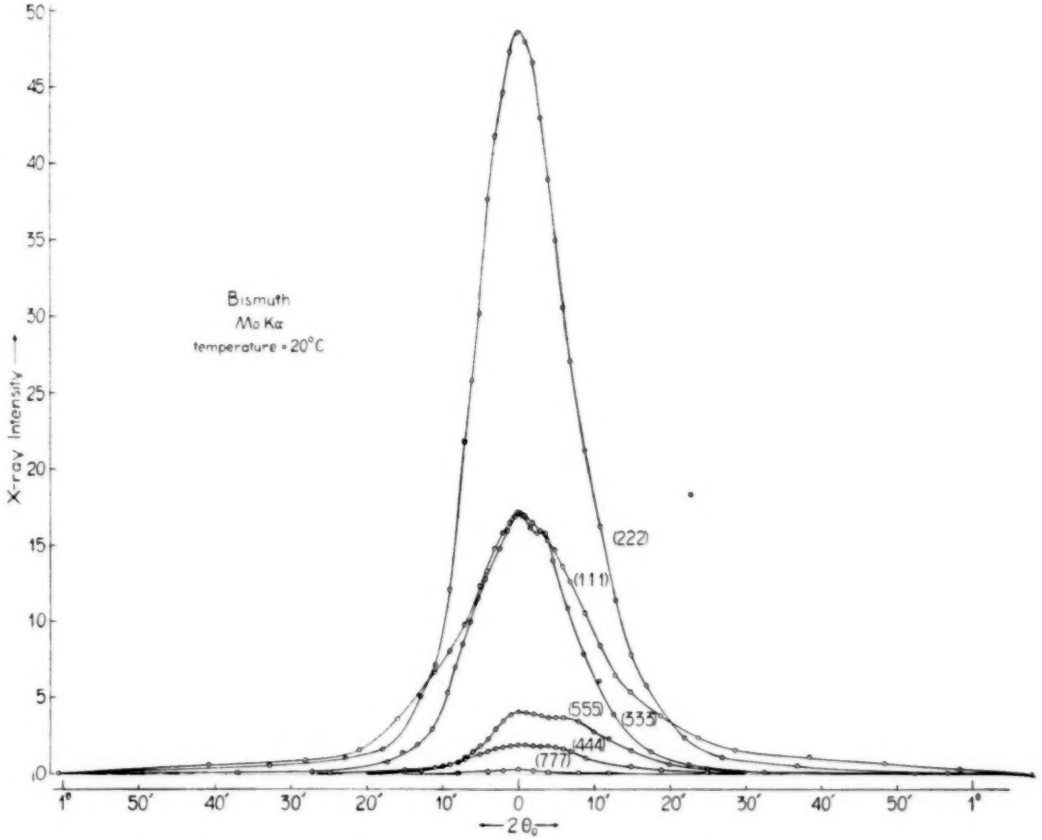

Fig. 5a represents the 1st to 7th order of the $K\alpha$-doublet reflection on (111) shifted in such a manner that the maximum intensities of $K\alpha_1$ coincide at 0°. Abscissa in minutes.

The Quantitative Determination of the Parameter

In order to determine the absolute value of the parameter an exceptionally good cleavage of a "normal" crystal (111) was selected, which permitted one to measure the intensity curves very accurately up to the seventh order. Fig. 5 a, b represent the diagrams obtained from which the quality of the crystal can be judged. Fig. 5a gives the first seven orders shifted in such a manner that the maximum intensities of $K\alpha_1$ coincide at 0°, the abscissa is shifted to the level of the background. Fig. 5b shows the last four orders

again for a tenfold ordinate scale. It appears that the 6th order is zero within experimental error.

In order to compute the parameter the following procedure has been used. First the atomic structure factor, defined as

$$F = z \int_{-\infty}^{+\infty} p(z) \cos (4\pi z/\lambda \cdot \sin \theta) dz^{10}$$

has to be known. The function $p(z)$—the probability of an electron being between the planes the distances of which from the atomic midplanes are respectively z and $z+dz$, has been recently calculated by Pauling and Sherman[11] assuming hydrogen-like eigenfunctions for each of the component electrons of

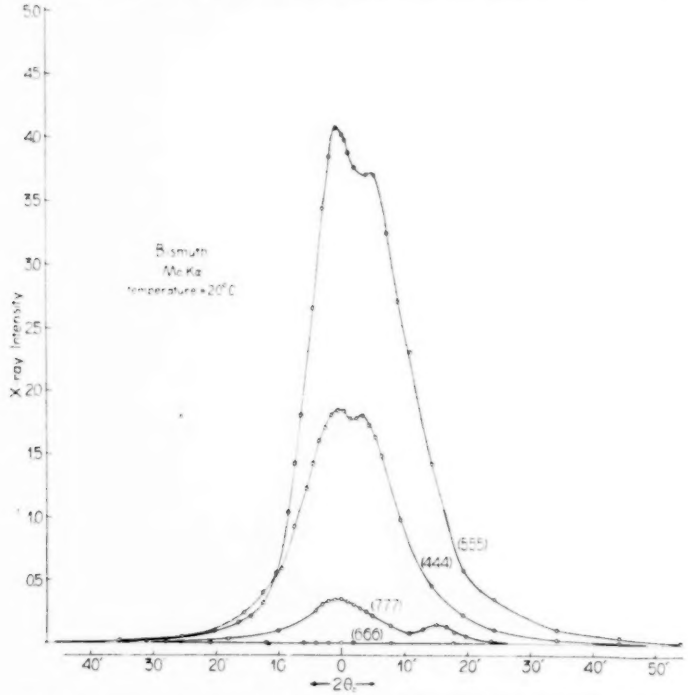

Fig. 5b represents the 4th–7th of Fig. 5a on a tenfold ordinate scale.

the atom. Those values of F are shown in Table IV for values of $\sin \theta/\lambda$ from 0 to 1.

TABLE IV.

$\sin \theta/\lambda$	F	$\sin \theta/\lambda$	F
0.0	83.00	0.6	41.55
0.1	75.53	0.7	37.15
0.2	65.83	0.8	33.03
0.3	56.92	0.9	29.24
0.4	50.59	1.0	26.18
0.5	46.11		

[10] A. H. Compton, X-rays and Electrons, page 122.

[11] L. Pauling and J. Sherman, Zeits. f. Kristallographie **8,** 1 (1932).

From these values the structure factor for the Bragg angles of the orders in question here interpolated. F being known, the relative values of the integrated reflected intensities at zero temperature J_{r_0} can be calculated for different orders by the relation:

$$J_{r_0} \sim J_i \, F^2 \left(\frac{1 + \cos^2 2\theta}{\sin 2\theta} \right)$$

where J_i is the intensity of the incident radiation. The exceedingly small depth in which Mo $K\alpha$ radiation can penetrate into Bi ($Z=83$) seemed to justify the assumption of the absence of primary extinction and therefore of the direct dependence of J_{r_0} on F^2. The numerical values for J_{r_0} are given in Table V. The existence of a displacement b of atoms along (111) defined by

TABLE V.

n	θ	$\sin\theta/\lambda$	F	$F^2 \frac{1+\cos^2 2\theta}{\sin 2\theta} 10^{-3}$
1	5.105°	0.1226	73.20	59.50
2	10.335°	0.2535	60.80	19.67
3	15.612°	0.3800	51.60	8.880
4	21.025°	0.5070	45.80	4.850
5	26.647°	0.6330	40.00	2.710
6	32.560°	0.7600	34.60	1.557
7	38.900°	0.8862	29.70	0.932

the parameter b/d, introduces a certain phase relation which superimposes another function on $J_r = f(F, \theta)$. The modified integrated intensity at zero temperature is given by:

$$J_{rm_0} = J_{r_0}(2(1 + \cos 2\pi n b/d))^{1/2}$$

Fig. 6 shows the curves of relative intensities for the first seven orders as function of parameter values from 0.35 to 0.5.

In order to compare the measured intensities from Fig. 5 a and b, which had been integrated by a planimeter, it was necessary to find the Debye factor:

$$D = e^{-B\sin^2\theta}$$

for the (111) direction in Bi. This was done by measurements at liquid air temperature[12] from where the $J_r = f(T)$—curves were extrapolated to 0° K.

Table VI gives the values of intensities, measured at room temperature (J_{296}) and the reduced values (J_0).

TABLE VI.

Order	$1/\sin^2\theta$	J_2^{92}	J_0
1	125.0	2205	2308
2	31.1	4400	5235
3	13.85	1590	2355
4	7.78	190	383
5	4.97	408	1217
6	3.45	0	0
7	2.54	32.0	272

[12] A. Goetz and R. C. Hergenrother, Phys. Rev. in press.

In Fig. 7 the relative intensities are plotted on a logarithmic scale against the parameter similar to Fig. 6, the parameter is marked where the intensities values determined by experiment fit best, thus b/d comes out to be 0.409 ± 0.0007. The agreement is obviously good except for the first order where the nature of the crystal surface plays too large a rôle and for the second order where the value of the reflected intensity is largest and hence the effect of secondary extinction is greatest. Since a correction for this could not be made this deviation had to be expected.

Since several definitions of the parameter are used by different authors,

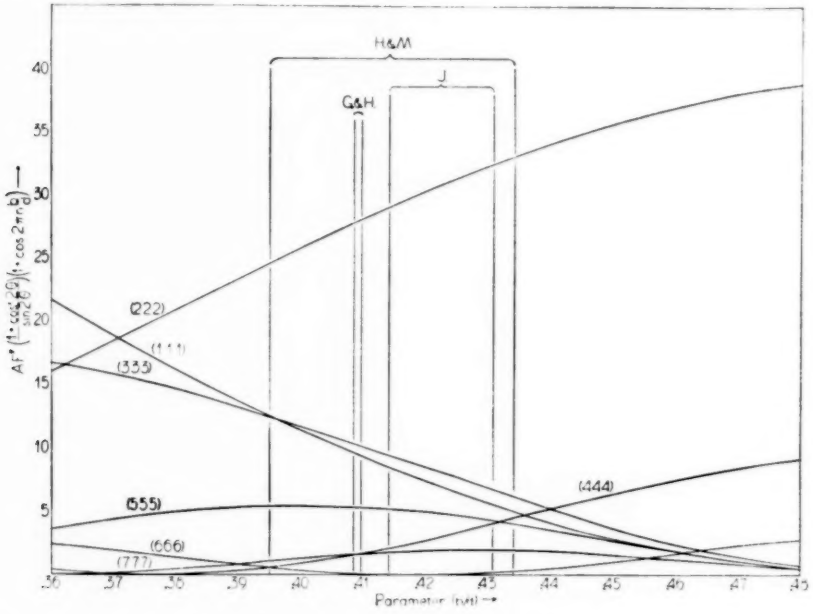

Fig. 6. Curves of relative intensities for the first seven orders as function of parameter values (b/d). The curly brackets include the possible parameter values of Hassel and Mark (H and M), of James (J), and of the authors (G and H).

the comparison can be made as follows: Using James'[7] definition our value is:

$$b' = \tfrac{1}{2}d - b = 0.359\text{A} \pm 0.0006\text{A}$$

which is slightly larger than the value of this author $= 0.307 \pm 0.033$A. Hassel and Mark express their parameter in terms of $c = 3 \cdot d = 11.85$A as 0.0285 ± 0.0065. Our value is then:

$$b'' = b'/c = \tfrac{1}{3}(0.5 - b/d) = 0.0303 \pm 0.00005$$

which is well within the order of their experimental error.

The mutual relation of the different values are given in Fig. 6, where the curly brackets include the possible parameter values of Hassel and Mark (H&M) of James (J) and of the authors (G&H).

Discussion and Summary

A precision measurement of the interplanar spacing along $(11\bar{1})$ of the bismuth lattice with a modified Siegbahn method has been performed on crystals formed within and outside of a magnetic field, in order to investigate the cause of the differences in density, thermal e.m.f. and electric conductivity, which have been described previously. It was found that the spacing in the most sensitive direction of the crystal does not show a change within 0.003 percent accuracy. No change could be observed either along $(11\bar{1})$, although the accuracy amounts only to 0.5 percent due to the difficulty of obtaining good planes.

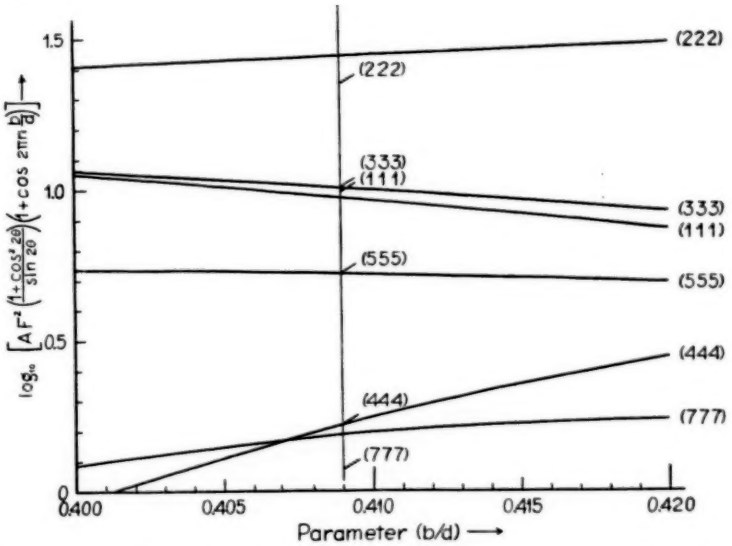

Fig. 7. Curves of relative intensities plotted on a logarithmic scale against the parameter (b/d). The observed J_0 values of the authors are indicated at the best fitting position $(b/d = 0.409)$. The sixth order (666) which cannot be shown on the log scale, agrees within experimental error. The seventh order agrees within the estimated experimental error.

The spacing along (111) was found to be $3.9453\text{A} \pm 0.0002\text{A}$ at 23°C. The absence of a change in spacing in spite of the observed change in macroscopic density supports the previous view that the crystallization within a magnetic field affects only the mosaic structure of the crystal and not the lattice, by causing differences in the packing of the mosaic units.

The large thermoelectric effects made it desirable to investigate whether or not a difference in the lattice parameter could be found between the two types of crystals. A new photographic recording method was used and in addition an ionization method of high sensitivity which permitted the quantitative observation up to the 7th order. Pauling and Sherman's wave-mechanical determination of the structure factor being used and the Debye factor being measured between room and liquid air temperature, the parameter along (111) was computed to $0.359 \pm 0.0006\text{A}$. No change of the parameter between "magnetic" and "normal" crystals could be found within a possible

error estimated to less than 1 percent. This proves that the conditions for the existence and the distribution of the "free" electrons as given by the geometric configuration of the atoms in the lattice stay unchanged. The large thermoelectric volume effects can therefore only be caused by variations in the lattice, the period of which must be many times larger than the size of an elementary rhomboheder in order to be undetectable by x-ray analysis. This implies the assumption of very large "mean free paths" for the "free" electrons responsible for the thermoelectric effect. This conclusion is in agreement with the well-known large structure sensitivity of these effects in bismuth crystals.

In conclusion the authors wish to express their indebtedness to Dr. J. W. M. DuMond of this Institute for a number of technical advices and to Dr. L. Pauling and Mr. J. Sherman of the Gates Chemical Laboratory for the kind permission to use the results of their calculated values of the atomic structure factor of Bi before publication.

NOTE ON THE SATELLITES ACCOMPANYING THE X-RAY DIAGRAM LINE $K\beta_1$

By F. R. Hirsh, Jr.
Department of Physics, Cornell University

(Received March 3, 1932)

Abstract

Deodhar has assumed the non-diagram lines $K\beta_2$, $K\beta'''$ and $K\eta$ to be satellites of the diagram line $K\alpha_1$, and not satellites of the diagram line $K\beta_1$ which they seem to accompany. His conclusion is due to the linearity of the satellite semi-Moseley graph, which results when $K\alpha_1$, and not $K\beta_1$, is considered to be the parent line of $K\beta_3$, $K\beta'''$ and $K\eta$. In the present note, the linearity of satellite data is shown to be entirely fortuitous. A study of the errors involved supports this conclusion. If we assume that these three non-diagram lines are really satellites of $K\beta_1$, the $\Delta\nu/R$ values which result, are of the same order of magnitude as the $\Delta\nu/R$ values for other satellites. The conclusion is that $K\beta_3$, $K\beta'''$ and $K\eta$ are probably satellites of the diagram line $K\beta_1$.

IN A recent paper,[1] G. B. Deodhar has shown that if we regard the diagram line $K\beta_1$ as the parent line of the satellites called by him $K\beta_3$, $K\beta'''$ and $K\eta$, confusion apparently results. We are then confronted with one short

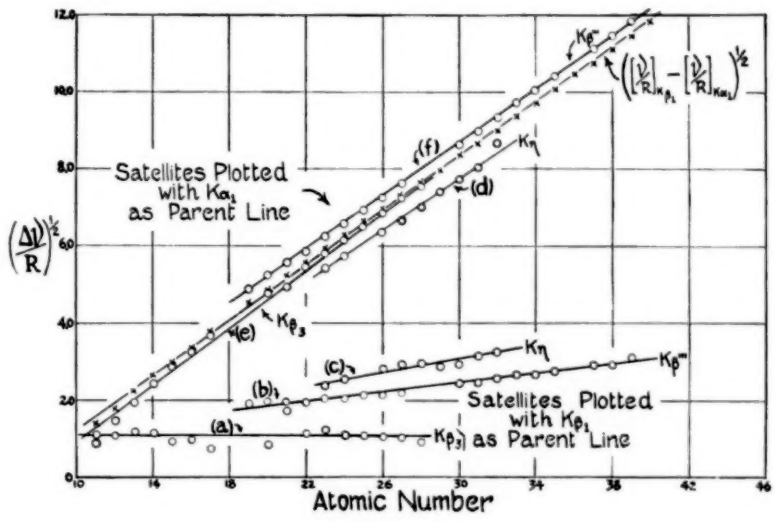

Fig. 1.

wave-length satellite, and two long wave-length satellites of the diagram line $K\beta_1$. Moreover, the square root of the difference in ν/R values of $K\beta_3$, $K\beta'''$ and $K\eta$ as compared with $K\beta_1$, is not a linear function of atomic number. The data vary considerably from linearity, (See Fig. 1).

[1] G. B. Deodhar, Proc. Roy. Soc. **131**, 476 (1931).

Up to the present time, this satellite Moseley law, discovered by F. K. Richtmyer,[2] has been assumed to be valid, if we always consider the nearest strong diagram line as the parent line of the satellites concerned. Deodhar, however, judges from the scattered $(\Delta\nu/R)^{1/2}$ values, and their apparent lack of linearity, Fig. 1, Curves *a*, *b* and *c*), that the non-diagram lines $K\beta_2$, $K\beta'''$ and $K\eta$ are not satellites of the diagram line $K\beta_1$. Moreover, when he secures values of $(\Delta\nu/R)^{1/2}$ for these three satellites with respect to the diagram line $K\alpha_1$, the behavior of these data is remarkably simple. The points lie quite accurately on a straight line. Assuming $K\alpha_1$ to be the parent line, we find that all three non-diagram lines might be considered as short wave-length satellites.

The purpose of this note is to show that, notwithstanding this apparently simplifying relation noted by Deodhar, the three non-diagram lines concerned are probably satellites of the diagram line $K\beta_1$. The reasons for this conclusion are:

(1) The lines $K\beta_2$, $K\beta'''$ and $K\eta$ occur in the x-ray spectrum in the same region as the diagram line $K\beta_1$.

(2) If the lines $K\beta_2$, $K\beta'''$ and $K\eta$ are associated with $K\alpha_1$ as parent line, the resulting $\Delta\nu/R$ values, in ν/R units, are entirely discordant with the $\Delta\nu/R$ values for other satellites. The $\Delta\nu/R$ values for $K\beta_2$, $K\beta'''$ and $K\eta$, referred to $K\alpha_1$, are entirely too great, ranging between 1.0 and 140.0, in ν/R units. All other satellites have $\Delta\nu/R$ values ranging between 0.25 and 5.0. If we assume that $K\beta_2$, $K\beta'''$ and $K\eta$ are satellites of $K\beta_1$, the resulting $\Delta\nu/R$ values lie between 1.0 and 10.0, in better agreement with the values for other satellites.

(3) A consideration of the assumptions made by Deodhar, and a study of the errors involved in the data, show without question, that when we consider $K\alpha_1$ as the parent line of $K\beta_2$, $K\beta'''$ and $K\eta$, the resulting linearity of $(\Delta\nu/R)^{1/2}$ values, in the semi-Moseley graph, is an entirely fortuitous result. The errors are minimized for purely mathematical reasons, when we assume that $K\alpha_1$, and not $K\beta_1$, is the parent line of the three satellites in question.

In Fig. 1, curves *a*, *b* and *c*, the data, given by Deodhar in Tables XII, XIII and XIV in his article,[1] are plotted with $K\beta_1$ considered as the parent line for $K\beta_2$, $K\beta'''$ and $K\eta$. There is obviously considerable deviation from the average straight line. Also in Fig. 1, curves *d*, *e* and *f*, are shown Deodhar's data for these three non-diagram lines plotted with $K\alpha_1$ as the parent line. The remarkable improvement in linearity is obvious. The crosses in Fig. 1 show a plot of values of

$$[\Delta_{K\beta_1}^{K\alpha_1}]^{1/2} = \left[\left(\frac{\nu}{R}\right)_{K\beta_1} - \left(\frac{\nu}{R}\right)_{K\alpha_1}\right]^{1/2},$$

for the two diagram lines $K\alpha_1$ and $K\beta_1$. There is seen to result a linear function of atomic number, from atomic number 11 to atomic number 40. Especially to be noted is the fact that this same value of $\Delta\nu/R$, namely $\Delta_{K\beta_1}^{K\alpha_1}$ for the particular atomic number involved, is included in each value of $(\Delta\nu/R)^{1/2}$

[2] F. K. Richtmyer, Phil. Mag. **6**, 64 (1928).

which is calculated with $K\alpha_1$ as parent line. This last point is best demonstrated by the following diagram, Fig. 2.
Considering $K\beta_1$ as parent line,—

$$\frac{\Delta\nu}{R} = \left(\frac{\nu}{R}\right)_{\text{SAT.}} - \left(\frac{\nu}{R}\right)_{K\beta_1}. \tag{1}$$

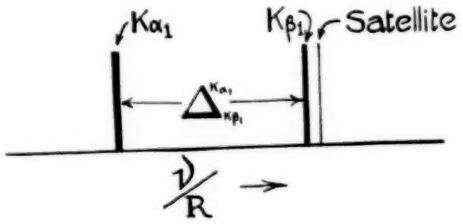

Fig. 2.

Considering $K\alpha_1$ as parent line,—

$$\frac{\Delta\nu}{R} = \left[\left(\frac{\nu}{R}\right)_{\text{SAT.}} - \left(\frac{\nu}{R}\right)_{K\beta_1}\right] + \Delta^{K\alpha_1}_{K\beta_1} \tag{2}$$

It is the presence of the large term, $\Delta^{K\alpha_1}_{K\beta_1}$, which decreases the errors in the satellite semi-Moseley graph data, when $(\Delta\nu/R)^{1/2}$ is plotted against atomic number with the non-diagram line in question treated as a satellite of $K\alpha_1$.

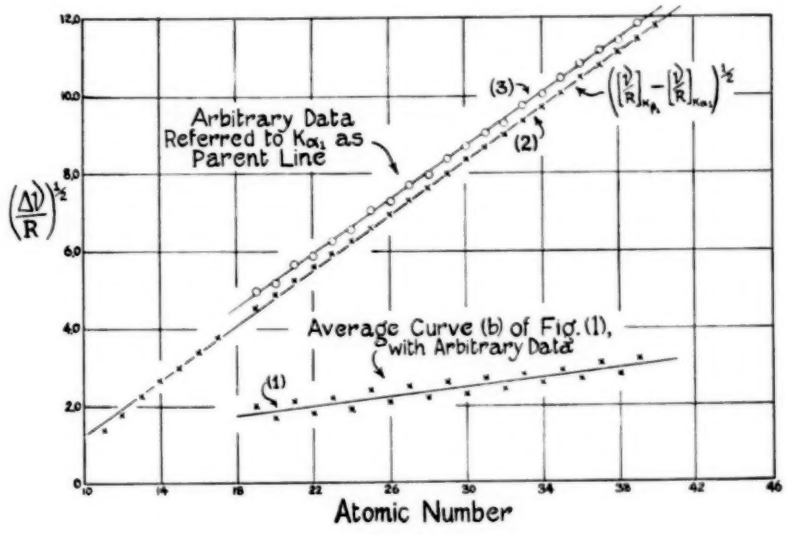

Fig. 3.

To support this contention, the average straight line derived from the data of the satellite $K\beta'''$ referred to $K\beta_1$ as parent line, (taken from Fig. 1), is transferred to Fig. 3, Curve (1). Around this line are grouped *a set of purely arbitrary data points, with an unusually large error of deviation from the average straight line.* In Fig. 3, values of $(\Delta^{K\alpha_1}_{K\beta_1})^{\frac{1}{2}}$ are also plotted as Curve (2). Above

them, in Fig. 3, Curve (3), are plotted values of $(\Delta\nu/R)^{1/2}$ for $K\beta'''$ referred to $K\alpha_1$ as parent line, which are obtained by using for $K\beta'''$ the arbitrary points of Curve (1). It will be seen that these perfectly arbitrary data, when treated according to Deodhar's assumptions, yield a set of data which obeys quite rigorously a linear relation, the deviation from the straight line being much reduced. The reason for this can easily be seen from the following brief discussion.

The three curves of Fig. 3 are represented diagramatically in Fig. 4. The notation of the three curves is obvious. Let us say that for atomic number Z, in Fig. (4), the ordinates of the three straight lines are respectively $h_{\alpha_1 S}$, $h_{\alpha_1\beta_1}$ and $h_{\beta_1 S}$. Let us assume that the data point for the curve at atomic number Z, has an absolute error, δ. We can calculate the absolute error for the data point of the curve $\alpha_1 S$, which is derived from the data point of curve

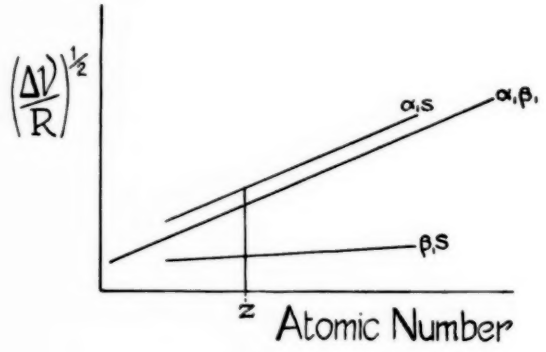

Fig. 4.

$\beta_1 S$, for atomic number Z. The ordinate for the data point for curve $\beta_1 S$ is represented by $(h_{\beta_1 S}+\delta)$. From Eq. (2), it is obvious that

$$h_{\alpha_1 S}^2 = h_{\alpha_1\beta_1}^2 + (h_{\beta_1 S} + \delta)^2.$$

Expanding this last equation, and neglecting the term δ^2, we get

$$h_{\alpha_1 S} = h_{\alpha_1\beta_1} + \tfrac{1}{2}\frac{h_{\beta_1 S}}{h_{\alpha_1\beta_1}} \times h_{\beta_1 S} + \frac{h_{\beta_1 S}}{h_{\alpha_1\beta_1}} \times \delta + \cdots . \qquad (3)$$

From Eq. (3), it is clear that the absolute error in the data point for the curve $\beta_1 S$ has been diminished by the ratio $h_{\beta_1 S}/h_{\alpha_1\beta_1}$, i.e., in the inverse ratio of the ordinates of the curves $\alpha_1\beta_1$ and $\beta_1 S$. From this it is evident that the farther away from the given satellites we choose a diagram line, the greater the reduction ratio will be for the original errors in the $(\Delta\nu/R)^{1/2}$ data. Obviously the square root of the difference in ν/R values of the new parent line, and the diagram line with which the satellites were formerly associated, should be a linear function of atomic number. Under these conditions, the non-diagram lines, referred to the new diagram line, will yield data points which apparently have a significant linearity. This linearity, however, has no physical significance, as the analysis of the errors has shown. This method

of deducing that $K\alpha_1$ is the parent line of the non-diagram lines $K\beta_3$, $K\beta'''$ and $K\eta$ is clearly at fault.

The result of this investigation seems to be that Deodhar's method of plotting data leads to erroneous conclusions. As far as all existing data warrants it, the conclusion must be that $K\beta_3$, $K\beta'''$ and $K\eta$ are most probably satellites of $K\beta_1$, and not satellites of $K\alpha_1$. It seems most plausible that the nearest strong diagram line to a group of non-diagram lines, is the parent line of that group of non-diagram lines. Exceptions to this rule may exist, but they have not yet been proven.

In conclusion, I wish to express my appreciation of the valuable criticism of Professor F. K. Richtmyer and Professor E. H. Kennard, which aided in the preparation of this note.

THE REFRACTIVE INDEX OF LIQUIDS FOR X-RAYS

By Scott W. Smith
University of Pittsburgh

(Received February 26, 1932)

Abstract

The refractive index for x-rays was measured for several liquids by the method of total reflection. A monochromatic beam ($\lambda = 1.537$A) was obtained by using a copper target in the x-ray tube and by selectively reflecting the beam from a calcite crystal. This beam was, in turn, reflected at a grazing angle from a thin film of the liquid being studied, spread on a plane glass mirror. As the index of refraction for x-rays is less than unity, the beam will be totally reflected at angles smaller than the critical angle, a sudden reduction occurring in the intensity of the reflected beam at the critical angle. Photographs, taken while the mirror was being rotated, showed such an edge. A discussion is given of the effects on the microphotometer curves of the slit width of the microphotometer, the width of the crystal slit, the rotation of the mirror and the finite width of the spectral line. From measurements of the curves, the values of δ, which is equal to $1-n$, were obtained and compared with the values calculated by means of the simplified form of the Drude-Lorentz dispersion formula. The results are.—

Substance	δ (experimental)	δ (theoretical)
Water	3.69×10^{-6}	3.53×10^{-6}
Butyl phthalate	3.57×10^{-6}	3.57×10^{-6}
Ethylene glycol	4.08×10^{-6}	3.88×10^{-6}
Triethanolamine	3.86×10^{-6}	3.91×10^{-6}
Glycerine	4.41×10^{-6}	4.34×10^{-6}
Glass	8.19×10^{-6}	.....

Introduction

MEASUREMENTS of the refractive index for x-rays of various solid materials have been made by several investigators. Stenström noticed that the wave-length of a spectral line, calculated by means of Bragg's formula gave values, for higher orders than the first, in disagreement with each other. This was attributed to the refraction of the x-rays in the crystal from which reflection was taking place. Using this deviation, values for the refractive indices of calcite, rock salt and gypsum were determined. A more direct method for measuring the refraction of x-rays seemed desirable. Since the use of a prism at minimum deviation, as in the case of ordinary light, had previously failed, Bergen Davis[1] and his co-workers suggested the use of unsymmetrical reflections from crystal wedges. A. H. Compton[2] was led, by the fact that the index of refraction for x-rays of most substances is less than unity, to suggest that its value could be determined from the critical angle of total reflection.

[1] B. Davis and H. M. Terrill, Proc. Nat. Acad. Sci. **8,** 357 (1922); B. Davis and R. V. Nardroff, Proc. Nat. Acad. Sci. **10,** 60 (1924); **10,** 384 (1924).

[2] H. Compton, Phys. Rev., **20,** 60 (1922); Phil. Mag. **45,** 1129 (1923).

The first two methods, the deviation from the Bragg relation and the use of the crystal wedge, are limited to solid crystals. The method of total reflection, however, is applicable to any substance which can present a smooth surface from which to reflect x-rays. It is only natural that solids should be chosen at first for the purpose of studying the refractive index by this method since they can, in general, be prepared with definite, plane, highly polished surfaces. However, it is evident that the method can be extended to the use of liquid films on solid mirror surfaces.

The Drude-Lorentz formula for dispersion may be applied to the case of x-rays in the simplified form, $n=1-\delta=1-Ne^2/2\pi m\nu^2$, provided the natural frequencies of the electrons in the atoms of the substance are sufficiently far removed from the frequency of the x-rays concerned. Here, n is the index of refraction, N is the number of electrons per unit volume, e and m refer to the charge and mass of the electron and ν is the frequency of the x-rays.

It is evident from the Drude-Lorentz formula, that the value of δ increases with an increase in the density of the material used as a reflecting surface and also with an increase in the wave-length of the x-rays. It has been pointed out, however, by Prins and others,[3] that increasing the wave-length of the x-rays or increasing the absorption coefficient of the material, greatly diminishes the sharpness with which the intensity of the reflection is reduced at the critical angle of total reflection. Since the absorption coefficient is roughly proportional to the density, the lighter substances should have a sharper critical edge.

In order to make an accurate determination of δ, it is desirable that δ be as large as possible, though it is equally important that the critical edge be sharp in order to locate accurately the critical angle of total reflection. A good compromise is obtained by using reasonably soft x-rays and a reflecting material with a small absorption coefficient. The use of organic liquids, therefore, should be especially appropriate since their absorption coefficients for x-rays are generally low. Futhermore, the use of the simplified form of the Drude-Lorentz formula is justified for the liquids and x-rays used in the work reported here. The copper $K\alpha_1$ line was used and the value of ν for this line ($\lambda=1.537$A.U.) is approximately 2×10^{18} per second. Oxygen was the heaviest constituent element in the liquids studied and the "frequency" of its K electrons is of the order of 0.1×10^{18} per second, making the ratio of the squares of these two frequencies only 1/400.

The author is aware of but one other measurement of the refractive index of liquids. Kellerman[4] reflected a beam of x-rays from a cylindrical surface on which was placed a thin film of the liquid to be studied. The photographs obtained show a direct and a reflected beam which fades out at the critical angle of total reflection. From a measurement of these photographs the critical angle was calculated. The geometry of the method is, however, somewhat

[3] J. A. Prins, Zeits. f. Physik **47,** 479 (1928); J. E. Henderson and E. R. Laird, Proc. Nat. Acad. Sci. **14,** 773 (1928); M. Schon, Zeits. f. Physik **58,** 165 (1929); E. Dershem, Phys. Rev. **34,** 1015 (1929); M. A. Valough, Jour. d. Physique et Rad. Ser. VII, **1,** 261 (1930).

[4] K. Kellerman, Ann. d. Physik, *Series 4,* **2,** 185 (1930).

involved and the author has, accordingly, endeavored to use a simpler method.

Method and Apparatus*

The method used here is essentially the same as the one commonly used in dealing with a solid mirror. A fine beam of monochromatic x-rays was reflected on to a photographic plate from the surface of a thin film of the liquid placed on a glass mirror. The films were so thin that there was no tendency, on tilting the mirror, for the liquid to flow but instead would rotate with the mirror, which was moved through only a few minutes of arc. As the surface was rotated, the intensity of the reflected beam was suddenly reduced at the critical angle of total reflection. If θ_c is the critical, grazing angle of incidence, we have, $\cos \theta_c/1 = n$, where n, the index of refraction, is slightly less than unity. It is usually convenient to use δ, where $n = 1 - \delta$ and therefore, $\delta = 1 - \cos \theta_c = 2 \sin^2 (\theta_c/2)$. Since θ_c is of the order of ten minutes of arc, the sine may be replaced by the angle and $\delta = \theta_c^2/2$.

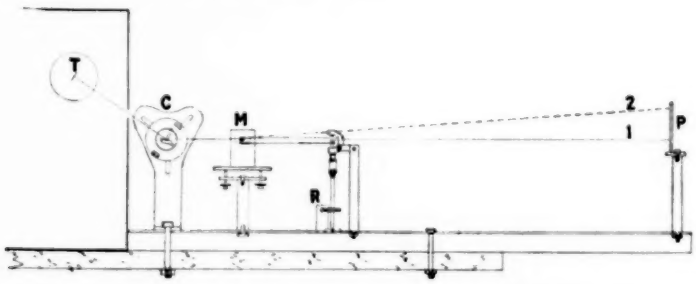

Fig. 1. *Schematic diagram of the apparatus.* T is a water-cooled, copper target, x-ray tube, operated at 36 kv and carrying a current of 25 m.a. Radiation from the target passed through a wide slit in the lead box housing the tube and fell on a calcite crystal C, which was provided with a Seemann wedge slit. The copper $K\alpha_1$ and $K\alpha_2$ lines were reflected horizontally from the crystal and passed through a thin cellophane window into a brass box housing the mirror, M. The position of M was so adjusted that the $K\alpha_2$ line always completely missed the mirror and a portion only of the $K\alpha_1$ line was reflected. The direct beam (1) and the reflected beam (2) left the box by another cellophane window and finally impinged upon a photographic plate P, which was placed 108.5 cms from the mirror. The mirror was rotated by means of a lever, one end of which was fastened to the axle of the mirror mounting and the other end rested on the micrometer caliper which was rotated very slowly by means of an electric clock motor.

In order to reduce the evaporation of the thin film of liquid on the surface of the mirror during an exposure, usually lasting for six to twelve hours, the mirror was mounted in a tight brass box measuring $6 \times 8 \times 10$ cms. A small dish containing some of the liquid was placed in the lower part of the box to assist in maintaining as nearly as possible, saturation of the vapor in the box. Some difficulty was experienced in getting the liquids to stick to the surface of the glass mirror, in a thin film. The method adopted was to clean the mirror thoroughly and then with a clean, rubber finger-stall on the finger, rub a small quantity of the liquid over the surface. The mirror was then placed in its mounting and installed in the brass box. In a short time an

* The method of experimenting was suggested by Dr. T. H. Osgood.

inspection through one of the cellophane windows would ascertain if the film had polished over uniformly or if it were necessary to repeat the process. It is to be noted, that in the case of the more viscous liquids used, even with the greatest care, it was not possible to obtain a perfectly smooth film over the entire surface of the mirror. The effect of this will be discussed along with the results.

The mirror mounting was supported in the box on two axles whose axis of rotation coincided with the surface of the mirror at its center. One axle fitted in a bearing in a fixed wall of the box and the other projected through a removable wall. On the outer end of this axle was clamped a rod, 26 cm long, for rotating the mirror. The end of this arm was raised by the micrometer caliper, driven by a small motor, so as to give the mirror a rotation of 3.4 minutes of arc per hour. For most of the photographs it was necessary to rotate the mirror three or more times through the same angle in order to obtain a suitable record on the plate.

The x-ray tube used was a copper target, water-cooled Coolidge type tube constructed in the laboratory and pumped continuously during operation. It was found that each time after a preliminary run of about half an hour, the tube would settle down to a very steady condition with fluctuations in the current of less than five percent. This would produce a corresponding fluctuation in the intensity of the reflected beam, but since the region was usually gone over several times, irregularities due to these fluctuations would tend to be smoothed out.

Microphotometer curves were obtained by means of a microphotometer constructed by Dr. Croup in the University laboratory.

Results

Before it was possible to make use of the microphotometer curves, of which Fig. 2 is a sample, it was necessary to study the effects of the width of the microphotometer slit, the wedge slit at the crystal, the rotation of the mirror, the finite width of the spectral line and the non-uniformity of the liquid film.

In order to determine the effect of the width of the microphotometer slit, a microphotogram was made of a photograph of a very sharp line drawn on a sheet of paper. A direct measurement of the width of the line on the photographic plate by means of a dividing engine and a measurement of the microphotogram of the line led to an effective slit width of 0.022 cm.

In the sample curve, Fig. 2, AC, the horizontal projection of the straight portion of the curve at the left edge of the direct beam is 0.41 mm, and DE, the horizontal projection of the straight portion in the region of the critical edge is 0.86 mm. In both cases this width is larger than 0.22 mm, the effective slit width of the microphotometer. Since such is the case, the effect of the slit here, as can be shown from simple considerations, is to round off the ends of the straight portion without displacing it or changing its slope. In the figure, a line is dotted in to show the expected microphotometer curve for a slit width of 0.22 mm, assuming the straight portion of the curve to extend

all the way from the maximum to the minimum ordinate. It is seen that this dotted curve, while it does not exactly fit the microphotogram, suggests that the assumption of a sharp bend at D is at least approximately correct. A better fit is noted at A and C.

Referring to the left portion of Fig. 3, two parallel monochromatic rays are drawn from the target T, which are limited by the Seemann slit A, and by reflection from the calcite crystal. It it is assumed, for the moment, that the crystal is perfect and the width of the spectral line is infinitesimal, each ray in the beam defined by these two limiting rays, in order to be reflected by the crystal, must be parallel to those rays.

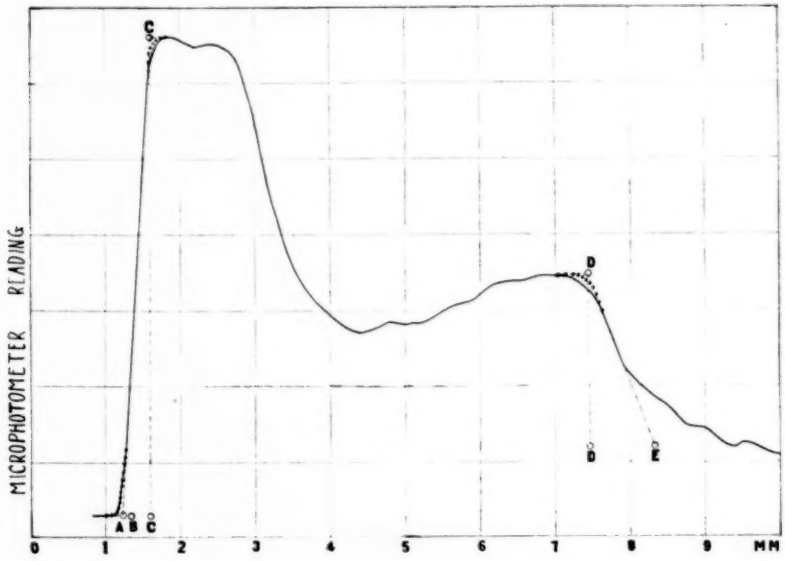

Fig. 2. *Sample microphotometer curve taken from a plate for ethylene glycol.* The broad, double hump, at the left of the figure is due to the direct beam. The right hand portion of this hump is due to the $K\alpha_2$ line which has entirely missed the mirror and the left hand portion of this hump is due to that part of the $K\alpha_1$ line which also missed the mirror. At the extreme left of the hump, however, is a region affected by that part of the $K\alpha_1$ line which was progressively cut out by the mirror as it was rotated. The part of the curve between the abscissas 6 and 8 mm, contains a hump which rises gently on the left and then falls off abruptly on the right, with a straight line as a portion of its right slope. Somewhere within the region of this hump lies the point corresponding to the critical edge. Microphotometer readings were taken every tenth of a millimeter.

From the geometry of the figure we have the relation, $CD = 2\ AB \cos\theta$. The distance AB, measured on a dividing engine, was 0.0142 cm. The value for θ for the $K\alpha_1$ line of copper is given by Siegbahn,[5] as 14° 41′ 46″. This leads to a value of 0.0137 cm for CD.

Suppose the mirror to be placed in the path of this parallel beam of width CD (as in Fig. 4) in such a way that the ray KL is reflected at the critical angle from the very edge of the mirror. In like fashion all other rays in the

[5] Siegbahn, The Spectroscopy of X-rays. Oxford University Press (1925).

beam between the limiting rays KL and IIJ will be reflected at the critical angle. The width of the beam KII will be less than that of CD, for the width of KII, calculated from the length of the mirror (4 cm) and the critical angle (for ethylene glycol, $\theta_c = 9' 34''$) is about 0.011 cm, which is less than the value of CD given above. Thus the mirror will always be reflecting over its entire length as long as it is not too near either edge of the $K\alpha_1$ line.

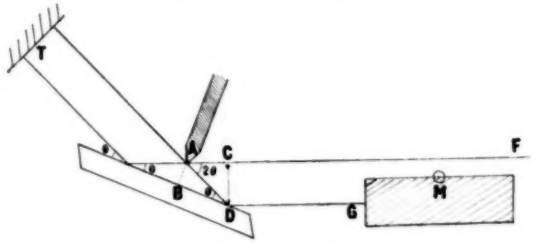

Fig. 3. *Diagram showing a beam from the anticathode T, crystal wedge slit and the mirror M.* The width of the spectral line is assumed zero and the crystal perfect.

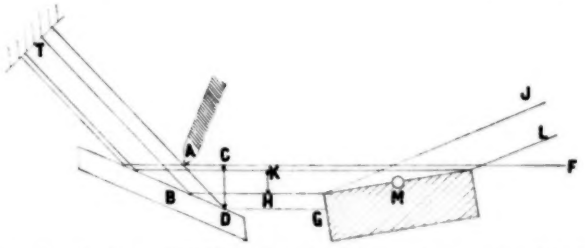

Fig. 4. *Diagram showing the beam KH limited by the mirror when at the critical angle.* The width of the spectral line is assumed zero and the crystal perfect.

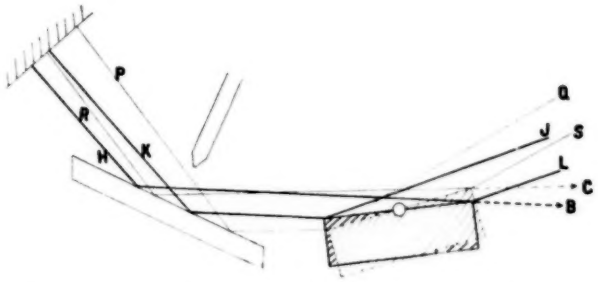

Fig. 5. *Diagram showing (1) the beam of shortest wave-length limited by the rays IIL and KJ, and (2) the beam of longest wave-length limited by the rays PQ and RS, reflected from the mirror in each case when the mirror is set at the critical angle.*

When the mirror is parallel to the beam, the width of the portion of the beam passing by the mirror is $CK+(1/2)KII$, or in the case of ethylene glycol, equal to $CK+0.0056$ cm, tending to give an image in the direct beam 0.0056 cm wider than when the mirror is at the critical angle. However, as the mirror is rotated from its parallel position to that of the critical angle, a portion of the direct beam will be progressively cut out, being reflected from the mirror. The result is that the edge of the direct beam is no longer sharp but shades off over a region 0.0056 cm wide. (AC Fig. 2).

It has been assumed so far that the radiation was strictly monochromatic. It is well-known that x-ray lines reflected from crystals, have finite widths due to the real width of the line, the finite resolving power and imperfection of the crystal, nonhomogeneity of the beam, etc. A. H. Compton[6] has found an angular width for the tungsten lines of about half a minute of arc. A measurement of the width of the copper $K\alpha_1$ line from a photograph taken with the mirror out of the way, resulted in a width at its half maximum of slightly more than two minutes. This larger value is to be expected for the longer wave-length copper line.

If we now consider the beam incident on the mirror as no longer composed of a single wave-length but having a definite, though small, range of wave-lengths, we may divide this range into a large number of beams, each of a single wave-length, and reflected at a slightly different angle from the crystal.

In Fig. 5, the beam *IIK* is the one of shortest wave-length which will be reflected from the mirror when it is set at the critical angle. The beam *PR* is the corresponding one of longest wave-length. Before the mirror has been rotated to the position for total reflection, the ray *IIL* will pass by the mirror and reach the photographic plate at B, Fig. 2, as a part of the direct beam. The region between *A* and *B* will be reached by the rays in the beam *IIK* which pass by the mirror as it is rotated from its horizontal position to the critical angle. This is only strictly true if the ray *KL* is limited by the wedge slit, as otherwise, rays of slightly shorter wave-length can pass the edge of the mirror.

Similarly, for the beam *PR* (Fig. 5), the ray *RS* will reach the point *C* (Fig. 2) before the mirror has rotated to the critical angle and the other rays in the beam which are able to miss the mirror will reach points a short distance to the left of *C*. Thus as the mirror is rotated the region *ABC* will be reached by rays which are progressively cut out. The point *B* then corresponds to the ray *IIL* of shortest wave-length and is reflected at the very edge of the mirror when it is set at the critical angle for this ray.

The critical point for the ray *IIL* will lie to the left of the critical points for the other rays of longer wave-length. There are two reasons for this, first, because the angle θ at the calcite crystal is smaller, and then, because the critical angle increases with the square of the wave-length. Thus the critical angle, will correspond to the point *D*, in Fig. 2, at the peak of the curve at the critical edge.

The reason that the curve falls off with a constant slope for some distance beyond the critical edge for the ray *CF* of shortest wave-length, is that the critical points for the rays of longer wave-length and the rays in other parts of the same monochromatic beam, fall farther and farther out beyond the one we are considering, and hence carry over some reflection beyond the first critical point. In fact the horizontal projection of the straight portion at the critical edge is approximately twice the horizontal projection of the straight portion at the left edge of the direct beam. This results from the

[6] A. H. Compton, Phys. Rev. **19**, 68 (1922).

fact that the reflected beam is rotated through twice the angle through which the mirror is rotated.

In order to save exposure time, the procedure usually was first to place the mirror parallel with the beam and then rotate it through a small angle in order to obtain the direct beam record on the plate. Next the mirror was turned to within a couple of minutes of arc of the critical angle and the slow rotation commenced. In consequence of this, there should appear on the plate, between the direct beam and the region in which the critical angle appears, a space which is unaffected. However, in Fig. 2, it can be observed that the minimum ordinate for the reflected radiation is somewhat greater than that for the left edge of the direct beam. This is due largely to the fact that the film of liquid was not strictly uniform. It is to be noted, however, that the curves obtained for the glass mirror, whose surface was very smooth, show the minimum ordinate for both the direct and reflected beam to be about the same. For the liquids in which difficulty was experienced in maintaining a uniform film, the difference in the minimum ordinate for the direct and reflected beam was greatest. This, however, does not obscure the critical edge since the bulk of the radiation is regularly reflected in the manner described earlier in this paper. The principal disadvantage lies in the reduction in intensity of the reflected beam and consequent necessity for longer exposure times.

It seems reasonable to assume that the point B in Fig. 2, corresponds to the position reached by the ray which is later reflected at the critical angle to D. The distance AB, calculated from the theoretical value of the critical

TABLE I. *Summary of results.*

Substance	Formula	Mol. Wt.	Density gm/cm³	Distance BD (cms)	θ Expt.	δ Expt. $\times 10^{-6}$	δ Calc. $\times 10^{-6}$
Water	H_2O	18.016	0.998	0.58 ± .02	9′20″	3.69 ± .20	3.53
Butyl Phthalate	$C_{16}H_{22}O_4$	278.22	1.167	0.569 ± .008	9′11″	3.57 ± .08	3.57
Ethylene Glycol	$C_2H_6O_2$	62.05	1.115	0.608 ± .009	9′49″	4.08 ± .09	3.88
Triethanolamine	$C_6H_{15}O_3N$	149.13	1.12	0.590 ± .007	9′33″	3.86 ± .09	3.91
Glycerine	$C_3H_8O_3$	92.06	1.26	0.632 ± .009	10′12″	4.41 ± .10	4.34
Glass*				0.861 ± .003	13′55″	8.19 ± .03	?

* The composition of the glass mirror was not definitely known.

Of the liquids measured by Kellerman, referred to earlier in the paper, glycerine was the only one also measured here. He also used the wave-length $\lambda = 1.537$A.U. and obtained for glycerine, $\delta = 3.8 \times 10^{-6}$, which is somewhat lower than the value obtained here.

angle and the length of the mirror, varied from 0.0053 cm for water to 0.0059 cm for glycerine. Since the points B and D correspond to the direct and reflected ray at the edge of the mirror, the distance from the photographic plate was measured to the edge of the mirror. This measurement was made by placing a long steel rod against the plate and making a scratch on the rod at the point corresponding to the edge of the mirror. The rod was then measured on a good metal metric scale, obtaining 108.42 cm as the distance. The distance measured on the microphotograms have a probable error of about 0.08 mm for all the liquids except water, for which the plates obtained were not as satisfactory as in the case of the other less volatile liquids. The experimental values for δ were calculated from the distance BD measured on the plate by means of the microphotometer curves, from five to eight curves being

obtained for different positions on each plate. The error is probably less than two percent. It is further seen, in the following table, that the values calculated from the Drude-Lorentz dispersion formula give values for δ in agreement within the experimental error.

Conclusion

The method of total reflection for measuring the refractive index for x-rays gives values for liquids comparable in accuracy with those for solids. The method is applicable to any liquid which can be kept in the form of a film on a smooth solid surface. By modifying the apparatus so as to keep the solid mirror at a lower temperature than the surrounding atmosphere, it should be possible to extend this method to more highly volatile liquids.

The author wishes to acknowledge his indebtedness to Dr. T. H. Osgood for his guidance in this problem, also to Dr. A. G. Worthing and Dr. E. Hutchisson for their interest in the preparation of this paper.

A SEARCH FOR PREFERENTIALLY DIRECTED ELECTRON VELOCITIES IN CRYSTALLINE GRAPHITE WITH THE MULTICRYSTAL SPECTROGRAPH

By Jesse W. M. DuMond, Harry A. Kirkpatrick and Lucas Alden
Norman Bridge Laboratory of Physics
California Institute of Technology, Pasadena, Calif.

(Received March 8, 1932)

Abstract

The author's theory of the broadening of the Compton line as a Doppler effect of electron velocities is briefly reviewed and it is pointed out that only the *component velocity* along a direction which nearly bisects the angle between primary and scattered x-ray beams should be effective in broadening the line. Crystalline Ceylon graphite possesses properties which lend hope to the belief that a class of weakly bound or structure electrons in this crystal might have momenta restricted uniquely to parallelism with one plane in the crystal: the (0001) plane. A composite scatterer was built up out of blocks consisting of the crystal flakes all orientated with their (0001) planes in mutual parallelism and the blocks in turn were so orientated that the normals to these planes bisected the angle formed by the primary and scattered x-ray beams. If the electron momenta are orientated parallel to the plane of the graphite flakes one should expect the contribution of such electrons to the shifted scattering to give a sharp line or peak superposed on the broader structure caused by the remaining isotropically distributed momenta. Details of the experimental set-up are described and the spectrum obtained from scattering by the Ceylon graphite scatterer is compared with the spectrum from an isotropic Acheson graphite scatterer. The breadth and structure of the shifted line proves to be quite identical in the two cases and the conclusion is drawn that if a class of electrons having selectively orientated momenta exists in the crystalline graphite it constitutes less than 5 percent of the total number. The bearing of this result on related questions is discussed.

Introduction

Interpretation of the breadth of the Compton modified line as a Doppler effect of electron motion

In A paper[1] published in this journal in 1929, the senior author proposed and analyzed the theory that the large observed spectral breadth of the Compton line can be regarded as a Doppler effect of the motions of electrons in the atoms and between the atoms of the solid scattering body. He has since then collected a large body of evidence[2,3,4] for the reality of this breadth, its structure and for the correctness of the Doppler interpretation.

The Doppler effect of the motion of a scattering particle changes the wave-length of the radiation scattered under a definite angle with the incident beam in a way that can be understood very simply by analogy with the classi-

[1] Jesse W. M. DuMond, Phys. Rev. **33,** 643 (1929).
[2] DuMond and Kirkpatrick, Phys. Rev. **37,** 136 (1931).
[3] Hoyt and DuMond, Phys. Rev. **37,** 1443 (1931).
[4] DuMond and Kirkpatrick, Phys. Rev. **38,** 1094 (1931).

cal Doppler effect of a moving mirror. Referring to Fig. 1, it is almost self evident that a motion of the mirror in the direction, y (or z) normal to x will leave the wave-length of the reflected beam unchanged, while a motion of the mirror along the direction, x, will impose a maximum Doppler effect on the wave-length of the reflected light. For a mirror moving with velocity, v, in any direction whatever only the component of v along x will be effective in modifying the wave-length of the reflected light.

The direction, x, will be for brevity called the longitudinal direction and the directions y, or z, normal to the axis, x, will be called transverse directions. If the change in wave-length relative to the wave-length itself is sufficiently small the direction, x, corresponding to maximum Doppler effect, very closely coincides with the bisector of the angle formed by the incident and reflected (or scattered) beams.

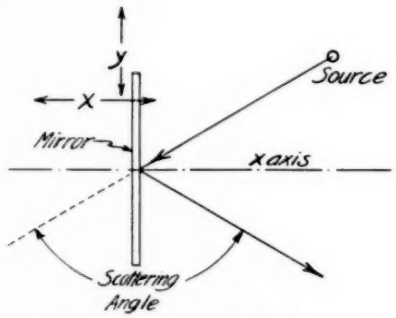

Fig. 1. Illustrating Doppler effect of a moving mirror.

It is evident that a given velocity of the mirror in Fig. 1 will modify the wave-length of the reflected beam more or less according as θ (the angle through which the reflection deviates the beam) is large or small. The effect will in fact be proportional to sin $(\theta/2)$ (to a non-relativistic approximation) being greatest if the beam is reflected directly back on itself as one would quite naturally expect.

All of the above statements can be taken over to the case of the scattering of x-rays by an electron if we replace the word mirror by the word electron. It must of course be remembered that while the mirror reflects the light almost uniquely in the specular direction the electron may scatter the quantum in any direction whatever. The statements we have made nevertheless apply to all those cases in which the electron does scatter the quantum in the particular direction, θ, chosen for observation.

Out of the multitude of electrons of a body scattering x-rays a certain fraction at any instant are scattering x-rays under a definite scattering angle, θ, (defined within certain narrow limits). What velocities should we expect these electrons to possess? We should expect them to have ordinarily a quite random distribution of velocities due to their quantized motions within the atoms and between the atoms *plus* a systematic velocity of recoil *away* from the x-rays. In accord with the law of conservation of momentum, this imparted recoil velocity will occur precisely in the direction x and will be pro-

portional to sin $(\theta/2)$ being greatest when the beam of radiation is scattered back along its original path.

We should expect, therefore, an initially sharp monochromatic wave to suffer a Doppler *shift* due to this imparted recoil velocity and a Doppler *broadening* due to the initial random velocities of the electrons. We should expect the shift to vary as $\sin^2(\theta/2)$ because the Doppler effect itself, as we have just pointed out, is proportional to sin $(\theta/2)$ and because the recoil velocity accountable for the shift is itself proportional to sin $(\theta/2)$. We should expect the broadening to be proportional only to the first power of sin $(\theta/2)$ because the initial random electron velocities do not depend on the scattering angle, θ. An accurate theoretical analysis of the problem verifies these qualitative statements and the results of experiment support the theory. In fact the shift above referred to is the well known Compton shift while the broadening is now well established by experiments of the senior author in conjunction with H. A. Kirkpatrick.[2,4]

That the Doppler explanation of the broadening is the true one can scarcely be doubted in view of the results of the senior author and H. A. Kirkpatrick which show that the broadening really does vary as sin $(\theta/2)$ *just as it should and further that the broadening varies with the primary wave-length used in the correct way.*

Effect of departure from isotropic momentum distribution

Up to the time of the present experiment the senior author has always assumed the directions of the initial electron velocities in and between the atoms of the scattering body to be quite random in orientation. This assumption is well justified for polycrystalline scatterers consisting of an immense number of individual crystals with completely random orientation. There remains, however, the possibility that in some monocrystals some of the electrons (especially the outer ones) may have directions of motion quite definitely related to the crystal structure. If a crystal exists in which a reasonably populous and isolated class of electrons have no component of motion whatever along some one axis in the crystal these electrons moving exclusively in directions parallel to a single plane then it should be possible to detect this fact by making this axis coincide with the x-axis of Fig. 1. We should then obtain not a broad Compton line but a narrow one as far as the scattering due to the anisotropically directed electrons is concerned. Even though this constituted but a small fraction of the entire shifted scattered radiation it should deform the now familiar shape of the Compton line by giving a more or less sharp peak at the center.

On the suggestion of Dr. A. Goetz of this Institute we decided to try Ceylon graphite crystals as the scattering body which might contain anisotropically directed electrons. Ceylon graphite cleaves much more readily along the (0001) plane than in any other direction. Also it is strongly diamagnetically anisotropic along an axis normal to the plane of easy cleavage. Both these facts encourage the hope that a class of electrons may perhaps possess velocities exclusively parallel to the plane of cleavage though they do not render it certain by any means.

The experiment seemed highly promising due to its bearing on several questions of interest in physics. As pointed out in a recent letter to the Editor of the Physical Review,[5] if the experiment should give a positive result, i.e., a narrow or peaked shifted line structure, this would be exceedingly, convincing corroboration of the correctness of the Doppler explanation of Compton line breadth.

A positive result might also explain the narrow line breadths reported by two other investigators[6,7] in sharp discord with our own observations and to some extent discordant with each other since these investigators might have used crystalline scatterers more or less fortuitiously orientated so as to reduce the longitudinal component of electron momentum. (Any departure from an isotropic momentum distribution might have a narrowing effect on the shifted line.)

Either a positive or negative result might throw considerable light on the behavior of the outer electrons in crystalline graphite. So little is known about the solid state that any information in this field is extremely valuable.

Finally, it should be emphasized that a negative result—i.e., a broad Compton line structure identical to that which is obtained from isotropic amorphous or highly polycrystalline scatterers is decidely *not* negative evidence for the author's Doppler interpretation of Compton line breadth as an effect of atomic electron momenta. *That theory is the only one yet proposed which satisfactorily explains the observed behavior of the breadth as a function of scattering angle and primary wave-length* and this is the strongest argument for its correctness. A negative result can only be reasonably interpreted as evidence that the outer electron momenta in the graphite crystals are not preferentially orientated as postulated above.

Theory of the Experiment

Simplified theory

Referring to Fig. 2 a simple form of the theory of scattering of x-rays by free or weakly bound electrons can be set up if, as a first approximation, we assume the change in wave-length due to scattering to be a small fraction of the wave-length itself and neglect the effect of relativity which is small. With this assumption the momentum imparted to the electron by the radiation is directed along x and of magnitude

$$2\frac{h\nu}{c}\sin(\theta/2). \tag{1}$$

The initial momentum of the electron is, mv, directed at angle ψ with x and the square of its final recoil momentum given by the parallelogram law is

$$p_2^2 = (mv)^2 + 4\left(\frac{h\nu}{c}\right)^2 \sin^2(\theta/2) + 4mv\frac{h\nu}{c}\sin(\theta/2)\cos\psi. \tag{2}$$

[5] Jesse W. M. DuMond, Phys. Rev. **39,** 166 (1932).
[6] J. A. Bearden, Phys. Rev. **36,** 791 (1930).
[7] Newell S. Gingrich, Phys. Rev. **36,** 1050 (1930).

If we divide this by $2m$ we have the energy imparted to the electron and hence lost by the radiation which can therefore be equated to $h(\nu_1-\nu_2)$. Transforming this in turn to a change in wave-length by multiplying through by $\lambda/h\nu$ (since $\nu_1-\nu_2 \ll \nu_1$) we obtain the shift

$$\lambda_2 - \lambda_1 = 2(h/mc)\sin^2(\theta/2) + 2\lambda\left(\frac{v}{c}\cos\psi\right)\sin(\theta/2). \tag{3}$$

The first term in the right member is the familiar "Compton shift" and the second term which may be either positive or negative accounts for the broadening of the line when we integrate over all possible values that v/c and $\cos\psi$ can assume in the multitude of scattering processes that occur. For a class of electrons of definite speed, v, and isotropically distributed as to spatial direction of motion the spectral breadth of the Compton line is evidently

$$\Delta\lambda = 4\beta\lambda\sin(\theta/2) \quad (\beta = v/c) \tag{4}$$

and the line, or rather band, should take the form of a rectangular spectral distribution bounded by sharp discontinuities as the senior author has shown

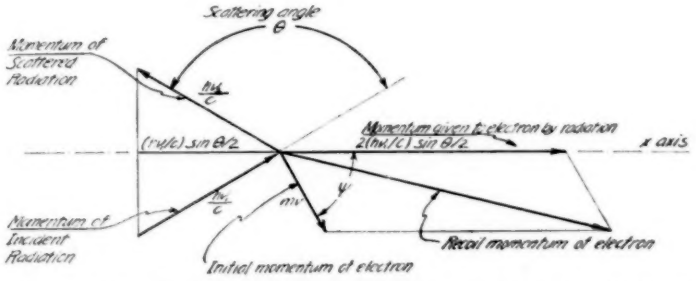

Fig. 2. Diagram of momentum vectors involved in scattering of radiation by a particle possessing initial momentum mv at angle, ψ_1 with an axis of reference bisecting the angle between the directions of incident and scattered rays. This is the simple approximation in which the change in wave-length is considered negligible in comparison to the wave-length itself. The angle, ψ, need not lie in the plane of the incident and scattered beams.

in previous papers.[1,2,4] (The observed shaded or continuous structure of the line can then be accounted for as an effect of the distributed values of v/c for all the electrons in the scatterer.)

Results of more elaborate theory

In a more elaborate treatment[2] the senior author has taken account of the fact that the change in wave-length is finite in comparison to the wave-length. The result obtained above is then only slightly modified and the breadth turns out to be for a random distribution of directions

$$\Delta\lambda = 4\beta\lambda^* \tag{5}$$

where $\lambda^* = \frac{1}{2}(\lambda_1^2+\lambda_c^2-2\lambda_1\lambda_c\cos\theta)^{1/2}$ (which becomes $\lambda\sin(\theta/2)$ when $\lambda_c=\lambda_1$); θ = scattering angle, λ_1 = primary wave-length, $\lambda_c = \lambda_1 + 2(h/mc)\sin^2(\theta/2)$ (shifted wave-length for initially stationary, electrons), $\beta = v/c$.

These results hold rigorously for free electrons and suffer from but slight error if the electron is weakly bound, the error being a fraction of the wavelength of the order of the binding energy divided by the quantum energy of the incident radiation.

The dependence of breadth, $\Delta\lambda$, on scattering angle and on primary wavelength, given in Eq. (5), has been tested by the senior author and H. A. Kirkpatrick using the multicrystal spectrograph and proves to be clearly verified experimentally.

The longitudinal axis in the more rigorous treatment just referred to does not quite bisect the angle between the incident and scattered radiation. It assumes in fact the direction of the vector difference between incident and scattered light quanta. This amounts, however, to a very slight difference indeed for molybdenum K radiation and hence has been neglected in the present experiment.

Effect of restriction of electron momenta to one plane

Reference to formula (3) shows that the broadening of the Compton line is to be ascribed to the *components* $(v/c)\cos\psi$ of all electron velocities resolved along the longitudinal axis, x, which very nearly bisects the angle between incident and scattered x-ray beams. If a scattering body could be found in which a reasonably large isolated class of electrons possessed no component velocity along some axis then the breadth and structure of the Compton line scattered by such a body would depend on the orientation of the body with respect to the longitudinal axis, x, bisecting the angle formed by the incident and scattered x-ray beams. With the axis of null velocity of the scattering body placed parallel to the axis, x, between the x-ray beams the contribution of the systematically directed electrons to the total scattering should be a narrow shifted line. One should expect to find this narrow line superposed on a broader diffuse line contributed by the scattering from the more tightly bound electrons whose directions of momentum would be less dependent on the lattice structure and hence practically isotropically distributed in space. With the axis of null velocity *normal* to the longitudinal axis between the x-ray beams one should expect a broader structure, as a contribution from the systematically directed electrons. In particular if these electrons had momenta isotropically distributed as to direction in the plane in which their vectors are restricted one should expect *two peaks* or maxima as the resulting spectral distribution when this plane is placed *parallel* to the longitudinal axis, x, between the x-ray beams. The perspective drawings in Fig. 3 illustrate the two cases just described together with the line structures one might expect to obtain from them.

Apparatus and Methods

Preparation of graphite blocks

The Ceylon graphite used in this experiment occurs in very small hexagonal flakes a few microns thick and from 0.01 to 0.03 mm across. The multicrystal spectrograph used for analyzing the scattered radiation requires a

large scattering body lying on a circular arc. For this experiment the circular scattering body was built up of fifty small blocks, one for each crystal of the multicrystal spectrograph, each block being composed of a multitude of the tiny flakes of Ceylon graphite stuck together with a slight quantity of gum Damar as a binder. A minimum amount of the gum was used so that the blocks were very fragile and had to be handled with the greatest care.

The technique of purifying the Ceylon graphite, preparing the blocks, and measuring their diamagnetic anisotropy was entirely worked out by Dr. A. Goetz and his collaborators, Dr. Faessler and Mr. Focke, and the authors wish here to express their deep appreciation for this very helpful cooperation. We wish also to thank Mr. L. Alden who purified a sufficiently large quantity of graphite for the fifty blocks and performed the tedious work of preparing the blocks and mounting them as a composite scattering body. In a recent letter[8] to the Editor of the Physical Review this technique has been described

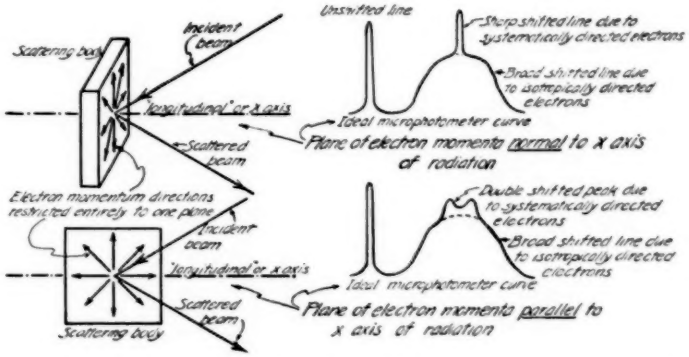
Scattering body
Incident beam
"Longitudinal" or x axis
Scattered beam
Electron momentum directions restricted entirely to one plane
Incident beam
"Longitudinal" or x axis
Scattered beam
Scattering body
Unshifted line
Sharp shifted line due to systematically directed electrons
Broad shifted line due to isotropically directed electrons
Ideal microphotometer curve
Plane of electron momenta normal to x axis of radiation
Double shifted peak due to systematically directed electrons
Broad shifted line due to isotropically directed electrons
Ideal microphotometer curve
Plane of electron momenta parallel to x axis of radiation

Fig. 3. Illustrating ideal Compton line shapes to be expected from a crystalline scatterer containing an isolated class of electrons with momentum vectors all parallel to one crystal plane and isotropically distributed therein.

at some length by Dr. Goetz, including the detection of the impurities (iron and quartz) in the graphite and the methods for removing them. The reader is referred to Dr. Goetz' article for this information.

The blocks were prepared by placing a suspension of the Ceylon graphite in a solution of 3 to 5 percent of the gum Damar in benzene. This suspension was held in small rectangular cellophane boxes open at the top. The cellophane box was placed in a horizontal magnetic field of a strong electromagnet. The suspension was allowed to settle until a sufficient quantity of graphite had accumulated at the bottom to form a block of the requisite thickness. By the combined action of the magnetic field and of the Bernoulli effect on the thin falling flakes these latter settled into horizontal positions with surprisingly good regularity. The alignment of the flakes was tested by examination of the block with a microscope both on its exterior surface and in some cases on the interior by cleavage of the block. An even better method of testing this alignment consisted in measuring the ratio of diamagnetism of the block

[8] Alexander Goetz, Alfred B. Focke and Alfred Faessler, Phys. Rev. **39**, 168 (1932).

normal and parallel to the general direction of the planes of the individual flakes. The ratio of diamagnetic anisotropy for the block as a whole can be taken as a measure of the perfection of crystal flake alignment. Even assuming the diamagnetic anisotropy ratio for individual flakes (comparing the directions normal and parallel to the flake) to be infinite, measurements of anisotropy ratios for the whole block indicate a departure of the flakes from parallelism of only a very few degrees. Blocks were obtained with anistropy ratios as high as 15.4. No blocks were used in the scatterer with anisotropy ratios lower than 9.2. It is probably safe to say that the mean deviation of the flakes from parallelism was of the order of three to five degrees of arc.

After the graphite had settled out of suspension in the cellophane holders these were placed in a rectangular hole in a block of brass into which they fitted. A rectangular brass block was then forced down through the open top

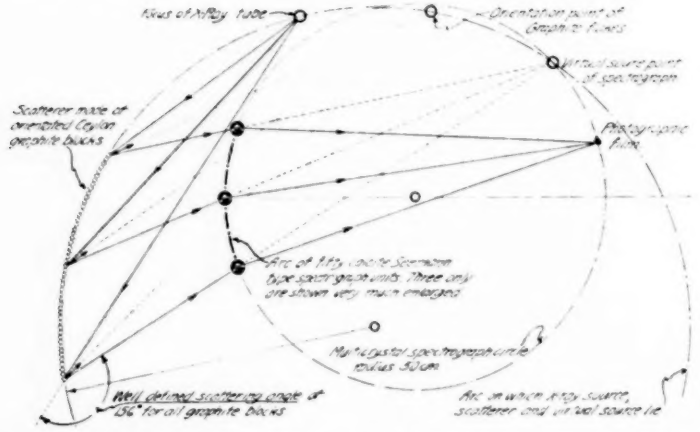
Focus of X-Ray tube
Orientation point of Graphite flakes
Virtual source point of spectrogram
Scatterer made of orientated Ceylon graphite blocks
Photographic film
Arc of X-Ray Crystal Spectrum (one such graph only. Three only are shown very much enlarged)
Multicrystal spectrograph circle radius 50 cm
Well defined scattering angle of 156° for all graphite blocks
Arc on which X-ray source, scatterer and virtual source lie

Fig. 4. Geometry of orientated Ceylon graphite block scatterer and multicrystal spectrograph fulfilling the conditions for (1) focusing the spectra from fifty crystals, (2) giving sharply defined scattering angle and (3) orientating the normals to the (0001) plane of the graphite flakes along the bisector of the angle between incident and scattered beams.

of the cellophane box and thus the deposited graphite was compressed and the excess fluid pressed out. After the block had dried the cellophane could be removed in most cases without damaging the block by exercising a little care.

Geometry and arrangement of scatterer and multicrystal spectrograph

The graphite blocks were of dimensions 50 mm×10 mm×8 to 10 mm, the planes of the flakes lying parallel to the 50 mm×10 mm face. The circular scattering body was built up of these blocks standing close together with their long dimensions vertical and each one so orientated that the normal to the plane of its flakes bisected the angle between its incident and scattered x-ray beam. Fig. 4 shows a plan view of the geometrical arrangement of the x-ray tube, the composite scatterer and the multicrystal spectrograph. The end of each block is shown and the small arrows indicate the normal to the planes of the graphite flakes in each block. These normals meet in a point on the large circular arc midway between the focal spot of the x-ray tube and

the virtual source point of the multicrystal spectrograph. The "focussing" principle of this spectrograph has been sufficiently explained in previous articles.[2,4,9] The present experiment differs from previous ones only in the nature of the scatterer consisting of individually orientated crystalline graphite blocks instead of a continuous curved strip of practically isotropic polycrystalline graphite. Fig. 5 is a photograph showing the x-ray tube, the scatterer of monocrystalline graphite blocks and the multicrystal spectrograph with its system of baffles in front of the crystals to reduce fogging of the film. It was of course necessary to be certain that each of the fifty small blocks was exactly in the proper position to furnish scattered radiation to its crystal in

Fig. 5. Photograph showing x-ray tube in cylindrical lead-rubber sheathed housing, scatterer consisting of fifty orientated Ceylon graphite blocks and multicrystal spectrograph with baffles to reduce fogging projecting in front.

the multicrystal spectrograph in the proper region of the spectrum, i.e., from about 800 X.U. to about 615 X.U. This was done by placing a strong light behind the photographic film-holder of the multicrystal spectrograph and stopping off everything except the region of the holder where the above mentioned spectral range was already known to appear. The light shining through this carefully defined opening is reflected from the fifty individual crystals of the multicrystal spectrograph along precisely the reverse path to be later followed by the scattered x-rays. The erect graphite blocks correctly spaced and orientated on a wooden board are then adjusted as a whole so that each reflected light beam from the fifty calcite crystals of the multicrystal spectrograph just covers the face of its respective graphite block. The entire geome-

[9] Jesse W. M. DuMond and Harry A. Kirkpatrick, Rev. Sci. Inst. **1**, 88 (1930).

try of the system is so simple that it is not at all difficult to position the blocks on their supporting board with the required accuracy by graphical methods. The geometrical elements were laid out to full scale on a large table and a thin board was cut with the proper saw tooth profile required to establish the proper positions and orientations of the bases of the graphite blocks. Two such saw tooth profiled boards attached to a base board formed a series of fifty depressions into which the bottoms of the graphite blocks fitted. A little cotton gently wedged into place behind each graphite block held it erect. These wooden base boards were covered with lead to prevent them from scattering x-rays. They were not in alignment to scatter radiation into the spectrograph.

The tube, scatterer and virtual source point in the spectrograph were es-

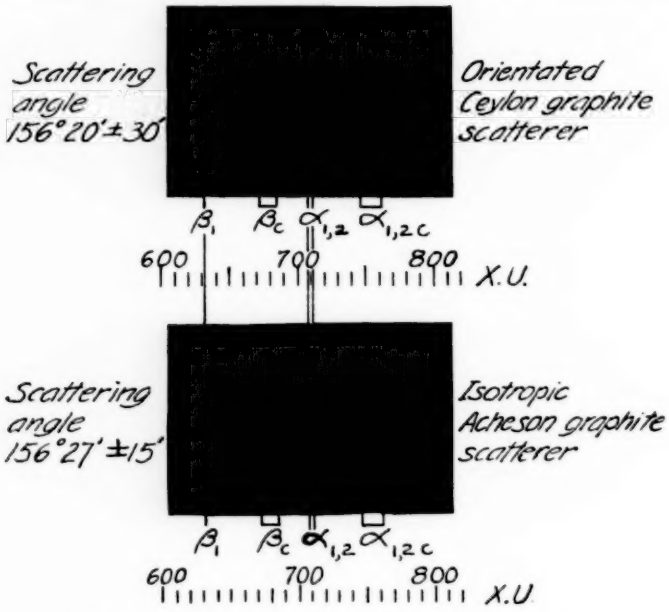

Fig. 6. Reproduction of spectrograms of molybdenum K radiation scattered at 156° 20′ ± 30′ from isotropic Acheson graphite and from orientated Ceylon graphite scatterers.

tablished on the same circular arc by means of the swinging radius arm with attached plumb line and the scattering angle was measured by observing the angle through which this arm must swing. In this respect the methods are entirely similar to those described in previous papers.[2,4,9]

Exposure conditions

An exposure of 990 hours was made with a molybdenum target x-ray tube running steadily night and day at 21 milliamperes current input and 50 kilovolts effective. The scattering angle was 156° 20′ ± 30′. At this large angle the tube must be placed so that the distance from the focal spot to the nearest point on the scatterer is about 70 cm and the distance to the farthest point is about 120 cms hence the necessity for such a prolonged exposure.

Eastman high speed duplitized "Diaphax" film was used without any in-

tensifying screen to avoid any possible falsification of the line-structure. The film was placed in a thin black paper envelope to avoid fogging by stray light in the spectrograph in the same manner as in previously described experiments.

RESULTS

A very fine clear exposure of the scattered spectrum was obtained—indeed one of the best to date. This exposure along with a similar exposure taken under identical conditions with a long continuous arc of isotropic Acheson graphite as a scatterer is reproduced in Fig. 6. *No essential difference in the structures of the shifted line for the two cases can be detected.* Neither do the microphotometer curves taken on the two spectrograms reveal any essential difference in the structures of the lines. Typical microphotometer curves are

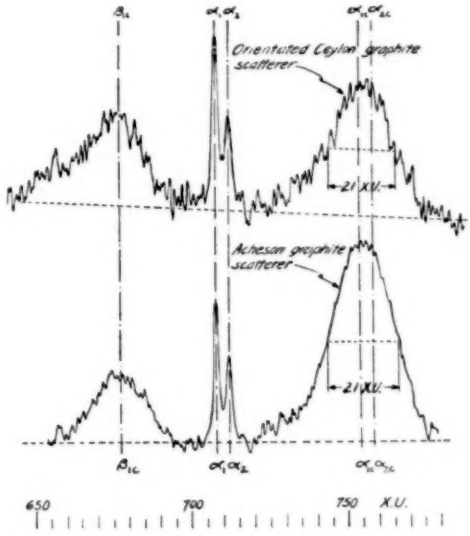

Fig. 7. Microphotometer curves taken from spectra of Fig. 6.

shown in Fig. 7. If anything the shifted line from the orientated graphite seems a little the broader and more diffuse of the two.

From a comparison of the areas under the shifted alpha line and under the unshifted beta line (which is clearly distinguishable on the original film) it is possible to form an estimate of what fraction of the total intensity of shifted alpha could have been detected if it had been present in the form of a narrow sharp line. The authors believe it is safe to say that a sharp line only 5 percent as intense as the total intensity of shifted alpha could easily have been detected if it were present. Since the scattering is proportional to the number of electrons it seems safe to conclude that if an isolated class of electrons exists in the crystalline graphite possessing negligible momentum components normal to the (0001) planes of the flakes this class must be less than 5 percent of all electrons or less than one to every three atoms. This estimate is highly conservative. The unshifted alpha doublet and the shifted beta line are more in-

tense relative to the shifted alpha line on the exposure with the orientated Ceylon graphite than they are on the exposures made with a continuous strip of Acheson graphite as a scatterer. This is probably caused by the nonuniform distribution of the scatterer over the range of wave-lengths of the spectrum on account of the blocks of which the scatterer consists, the different parts of the spectrum being scattered from different parts of the block.

There seems to be no reason to expect that any narrowing or peaking effect on the line would be found for other orientations of the graphite in view of the present result and as these experiments are expensive and very time consuming a further search for orientated velocities does not seem justified.

The film taken with isotropic Acheson graphite shows a region of slightly different density at one edge. This it will be noted, does not appear in the later film made with the orientated Ceylon graphite. This region of slightly greater blackening was accounted for by Dr. Kirkpatrick some months past as being due to a tiny leak in the construction of the lead spectrograph box which permitted the entry of a very small quantity of radiation scattered directly from the primary beam by the air outside the box, without suffering reflexion by the crystals. This was sufficient in the long period of exposure to cause this local fogging. This leak has since been completely eliminated. The possibility of falsification of the line structure by air scattering and subsequent selective reflection from the crystals is completely negligible since any such effect must be compared with scattering by the scatterer itself and it is an easy matter to compute the relative number of electrons in the region of air capable of scattering radiation to the crystals in comparison to the number of scattering electrons in the scatterer. This ratio is extremely small.

Conclusion

In the author's opinion the obvious conclusion to be drawn from this experiment is that the momenta of the electrons in the graphite crystals are isotropically directed in space, or if a class of electrons exists for which this is not true such a class constitutes a very small proportion of the total number (less than 5 percent).

It might have been supposed that the electrons responsible for the diamagnetic anisotropy of the graphite flakes executed flat orbits parallel to the (0001) plane of the flake. This also is the plane of easy cleavage and one might suppose that this easy cleavage could be explained by the flatness of such orbits and the consequent absence of any strong electron bonds due to electrons crossing these planes transversely. It must be noted, however, that even if this were the case it does not necessarily follow that these electrons will not possess *momenta* normal to the planes of easy cleavage. A rough picture of such a state of affairs can be visualized—the electron executing an orbit in a plane and at the same time performing small but rapid oscillations normal to the plane. Indeed the uncertainty principle leads us to expect that if an electron were sufficiently constrained in its *position* measured normal to some plane in which it might be supposed to execute an orbit, its spread in momentum along this same direction could become indefinitely large. The *shapes*

of the orbits and the distribution of electron momenta are thus not simply connected as one might naively expect from analogy with macroscopic particle dynamics.

The results of this experiment leave no opening for an explanation of the narrow shifted line structures reported by Gingrich and by Bearden.[6,7] Each of these men used the double crystal spectrometer. Dr. Hoyt and the senior author have, with the double crystal spectrometer, obtained good verification of their broad line structures in accord with the results obtained with the multicrystal spectrograph. Furthermore, P. A. Ross, by the use of very large polished calcites in a double spectrometer, has recently obtained very excellent spectral curves showing the broad Compton line also in complete agreement with our observations. In a letter to the author Ross states that he has also observed the variation in breadth with scattering angle first discovered by DuMond and Kirkpatrick. The breadth of Ross' shifted line is not an effect of polishing the crystals as he obtains narrow unshifted lines with them. The evidence for the broad Compton line and for DuMond's predicted dependence of that breadth on primary wave-length and scattering angle seems now very clear cut.

As stated above, the interpretation of the broadening of the Compton line as a Doppler effect of electron momenta in the scattering body is based on the observed behavior of the breadth with change of scattering angle and primary wave-length. The negative result of the experiment here reported leaves this interpretation still unchanged. Indeed it is the only interpretation so far offered which satisfactorily explains the observed experimental facts.

This investigation which has involved a great deal of expense has been carried out with the aid of funds from the Seeley W. Mudd X-Ray Research Fund and we take this opportunity to express our sincere appreciation for this financial aid.

ELASTIC REFLECTION OF ATOMS FROM CRYSTALS

By Clarence Zener*
Palmer Physical Laboratory, Princeton University

(Received February 8, 1932)

Abstract

The reflection of atoms by a crystal surface is examined by a modification of the Born collision method. The fraction of reflections in which all the normal coordinates of the crystal remain unaltered is found to be approximately

$$U = e^{-3\pi^2(m_A/m_B)(t_A/\Theta)^2(t_B/\Theta)^2/(1+4\pi d/\lambda)}.$$

m_A, m_B are the masses of the reflected atom A and of the lattice atoms, respectively. The energy of atom A is kt_A. The temperature of the lattice is t_B. Θ is the characteristic temperature of the lattice. d is the distance in which the mutual repulsion between atom A and lattice atoms falls to one e'th its value. λ is the de Broglie wave-length of atom A. For reflection of H atoms on LiF at room temperature, U is estimated to be 0.9. When Θ and t_A are comparable, those normal coordinates are found to be most readily excited which have the lowest frequency.

I. Introduction

RECENT experiments[1] indicate that the majority of light atoms bombarding a crystal may be reflected without an appreciable loss of energy. It is of interest to ask: what fraction of atoms will be perfectly elastically reflected? In the classical theory this fraction is infinitesimal. Some transfer of energy would always occur between lattice and bombarding atom. The situation is not so obvious from the standpoint of the quantum mechanics. Each normal coordinate of the lattice can absorb only quantised amounts of energy. But as the number of atoms in the lattice becomes infinite, the spectrum of the lattice approaches a continuum, so it is not evident that the quantum mechanics should hinder an exchange of energy. The following analysis will show however that this actually is the case.

It will be assumed that the only transitions of the crystal are those of the normal coordinates of the lattice. The excitation of electronic levels by the colliding atoms is considered negligible.

II. Formulation of Problem

Our problem will be stated in an idealised form. (See Figure 1.) Angular reflection difficulties will be avoided by confining the bombarding atom A to one dimension. Its line of motion is taken to be perpendicular to the surface of the lattice, and to pass through the equilibrium position of a definite surface atom B. The interaction energy is considered to be a function only of the distance R between the two atoms A and B, say $V(R)$.

* National Research Fellow.

[1] T. H. Johnson, Phys. Rev. **37**, 847 (1931); J. K. Roberts, Proc. Roy. Soc. **129**, 155 (1930) has discussed the similar implications of the experiments of Stern and his collaborators.

The crystal is to be of simple cubic type, with all the atoms of equal mass, m_B. There are N atoms on an edge, N^3 atoms in all. The three components of the displacement from equilibrium position of the atom with indices (lmn) shall be x_{lmn}, y_{lmn}, z_{lmn}.

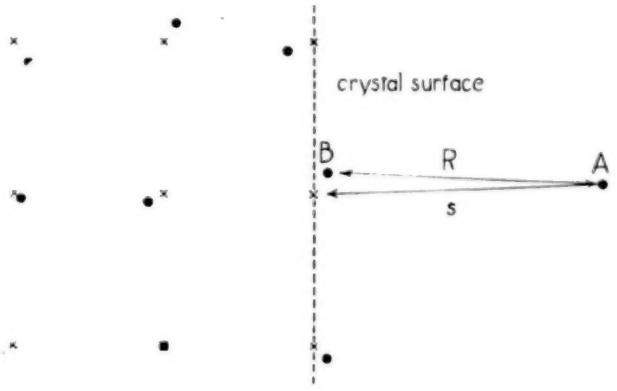

FIG. 1.

If in the lattice we assume the mutual energy of pairs of atoms to be a function only of their interatomic distance, the transformation[2]

$$\begin{aligned} x_{lmn} &= N^{-3/2} \sum_{\alpha\beta\gamma=-N/2}^{N/2} a_{\alpha\beta\gamma} e^{2\pi i(l\alpha+m\beta+n\gamma)/N} \\ y_{lmn} &= N^{-3/2} \sum_{\alpha\beta\gamma=-N/2}^{N/2} b_{\alpha\beta\gamma} e^{2\pi i(l\alpha+m\beta+n\gamma)/N} \\ z_{lmn} &= N^{-3/2} \sum_{\alpha\beta\gamma=-N/2}^{N/2} c_{\alpha\beta\gamma} e^{2\pi i(l\alpha+m\beta+n\gamma)/N} \end{aligned} \tag{1}$$

changes the Hamiltonian of crystal into nearly

$$\begin{aligned} H = \sum_{\alpha\beta\gamma} \{ & m_B/2(|\dot{a}_{\alpha\beta\gamma}|^2 + |\dot{b}_{\alpha\beta\gamma}|^2 + |\dot{c}_{\alpha\beta\gamma}|^2) \\ & + K(\sin^2 \pi\alpha/N + \sin^2 \pi\beta/N + \sin^2 \pi\gamma/N)(|a_{\alpha\beta\gamma}|^2 + |b_{\alpha\beta\gamma}|^2 + |c_{\alpha\beta\gamma}|^2)\}. \end{aligned}$$

The conditions

$$a_{-\alpha-\beta-\gamma} = a^*_{\alpha\beta\gamma}, \text{ etc.}$$

are necessary in order that the displacements be real. The terms that must be added to make this transformed Hamiltonian exact involve only the coordinates of the surface atoms. Our neglect of these terms is equivalent to assuming that the surface atoms behave as if the crystal were of infinite extent.

We are now able to write the wave equation of the system. The center of gravity of the total system is eliminated as usual. Let the mass of the free atom be m_A, and let its distance from the equilibrium position of the lattice

[2] M. Born and T. v. Karman, Phys. Zeits. **13**, 297 (1912).

atom B be s. Let $\xi_{x\alpha\beta\gamma}$, $\eta_{x\alpha\beta\gamma}$ be the real and imaginary parts of $a_{\alpha\beta\gamma}$, respectively. The wave equation is then

$$\Big\{ \sum_{\alpha\beta\gamma} \sum_{xyz} [(h^2/8\pi^2 m_B)(\partial^2/\partial\xi^2{}_{x\alpha\beta\gamma} + \partial^2/\partial\eta^2{}_{x\alpha\beta\gamma}) - \tfrac{1}{2} m_B(2\pi\nu_0)^2(2l/N)^2(\xi^2{}_{x\alpha\beta\gamma} + \eta^2{}_{x\alpha\beta\gamma})] + \frac{h^2}{8\pi^2 m_A}\frac{\partial^2}{\partial s^2} - V(R) + E\Big\}\psi = 0 \tag{2}$$

We have made the usual approximation of the Debye theory of specific heat that

$$K\{\sin^2(\pi\alpha/N) + \sin^2(\pi\beta/N) + \sin^2(\pi\gamma/N)\} = m_B(2\pi\nu_0)^2(2l/N)^2/2$$

where $l^2=\alpha^2+\beta^2+\gamma^2$ with $0<l\leqq N/2$.

Our problem may now be precisely defined. We are to find a solution of (2) which for large values of s satisfies the initial condition that when atom A is moving to the left, it has a definite energy kt_A, and the crystal is in a definite state corresponding to a temperature t_B. The solution is thus to have the asymptotic form

$$\psi \overset{s\to\infty}{=} e^{2\pi i(p_0/h)\cdot s}(p_0/m_A)^{-1/2}\phi_{\tau_0}(X) + \sum_\tau C_\tau e^{-2\pi i(p_\tau/h)\cdot s}(p_\tau/m_A)^{-1/2}\phi_\tau(X). \tag{3}$$

The variables of the crystal are denoted by X. $\Phi_\tau(X)$ is that normalized wave function of the crystal whose ensemble of quantum numbers is denoted by τ. If ϵ_τ is the eigenenergy of the crystal in state τ, then $E_\tau=E-\epsilon_\tau$ is the energy of atom A when the crystal is in state τ. p_τ refers to the corresponding momentum $(2m_AE_\tau)^{1/2}$.

The first term is normalized to represent a unit flux of atoms[3] to the left associated with the crystal in the initial state τ_0. The reflected atoms are represented by the summation. In order that the net flux of atoms across a given point vanish

$$\sum |C_\tau|^2 = 1. \tag{4}$$

In particular, the elastic reflection coefficient U is equal to $|C_{\tau_0}|^2$, or

$$U = 1 - \sum{}' |C_\tau|^2 = 1 - S \tag{5}$$

where the summation is over all states except τ_0, and is denoted by S.

III. Solution

A method previously used by the author[4] for studying the interchange of energy between molecules will be applied to calculate the coefficients of (3). In this method the first approximation to C_{τ_0} is unity, so the elastic reflection probability must be defined by (5) where the first approximations to the C_τ's are used.

[3] Cf. P. Dirac, Principles of Quantum Mechanics, Oxford Press (1930), p. 176.

[4] C. Zener, Phys. Rev. **37**, 556 (1931).

Let the x component of displacement of atom B from its position of equilibrium be denoted by x_B. Since atom A is confined to one dimension, the interaction $V(R)$ must be a function only of one dimension, so we set $R=s-x_B$. Then

$$V(R) = V(s) - x_B V'(s),$$

neglecting higher powers of x_B.

The perturbation scheme starts with the solution of (2) in which $V(s)$ is substituted for $V(R)$. The difference between $V(R)$ and $V(s)$, namely $-x_B \cdot V'(s)$, is taken as the perturbing term that causes transitions. The probability of the transition $\tau_0 \rightarrow \tau$ is given by[5]

$$\gamma^{\tau}_{\tau_0} = \left| (8\pi^2 m_A/h^2) \int \phi_{\tau_0} x_B \phi_{\tau} dX \int Q_{\tau_0} V'(s) Q_{\tau} ds \right|^2. \tag{6}$$

$Q_\tau(s)$ is that solution of

$$\{(h^2/8\pi^2 m_A) d^2/ds^2 - V(s) + E_\tau\} Q_\tau(s) = 0 \tag{7}$$

which has the asymptotic behavior

$$Q_\tau(s) \overset{s\to\infty}{=} \frac{\sin((2\pi/h) p_\tau s + \theta_\tau)}{(2\pi p_\tau/h)^{1/2}}.$$

In evaluating the first integral of (6) we replace the index B by the indices $(N/2\ m\ n)$, and use the transformation (1). We see that this integral vanishes except when one and only one normal coordinate changes its quantum number by ± 1. Let that normal coordinate which changes be $\xi_{x\alpha\beta\gamma}$ or $\eta_{x\alpha\beta\gamma}$, and let its initial quantum number be $n_{\alpha\beta\gamma}$. Then the value of $\left| \int \phi_{\tau_0} x_B \phi_\tau dX \right|^2$ averaged over all surface atoms, i.e., over m and n, is

$$N^{-3} \binom{n_{\alpha\beta\gamma} + 1}{n_{\alpha\beta\gamma}} h/(4\pi^2 m_B \nu_{\alpha\beta\gamma}).$$

The two quantities in the bracket are to be associated with the transitions ± 1. $\nu_{\alpha\beta\gamma}$ is related to the limiting frequency ν_0 by

$$\nu_{\alpha\beta\gamma} = \nu_0 2l/N.$$

A definite form for $V(s)$ must be assumed in the evaluation of the second integral of (6). This may be taken to be $Ce^{-s/d}$,[6] in which d has the magnitude of about $0.2\times 10.^{-8}$. Then $V'(s) = -V(s)/d$. However, with this potential Eq. (7) is difficult to solve. Hence we substitute a potential with which we can solve this equation. Such a potential is $A/(s-B)^2$, where the constants A and B are so adjusted that the two potentials join smoothly at that value of s for which $V(s) = E_{\tau_c}$, i.e., at the classical closest distance of approach. For ease in integration we shall still replace $V'(s)$ by $-V(s)/d$. These assumptions lead to[7]

[5] Reference 4, Eq. 13.

[6] Reference 4, Section II.

[7] Reference 4 (19). The E'' of this reference has been replaced by E_{τ_c} to facilitate future integrations.

$$\left|(8\pi^2 m_A/h^2)\int_0^\infty Q_{\tau_0}V'(s)Q_\tau ds\right|^2 = (2\pi^2 m_A/h^2)(\epsilon'/\epsilon)^{2d(2\pi p_{\tau_0}/h)}.$$

Here $\epsilon' = E_{\tau_0} - h\nu_{\alpha\beta\gamma}$, $\epsilon = E_{\tau_0}$, or $\epsilon' = E_{\tau_0}$, $\epsilon = E_{\tau_0} + h\nu_{\alpha\beta\gamma}$ according as $n_{\alpha\beta\gamma}$ changes by $+1$ or -1. $\lambda = p_{\tau_0}/h$ is the de Broglie wave-length of atom A before collision. Combining these integrations gives

$$\gamma_{\tau_0}{}^\tau = (\pi^2/2N^3)(m_A/m_B)(E_{\tau_0}/h\nu_{\alpha\beta\gamma})(n_{\alpha\beta\gamma} + \tfrac{1}{2} \pm \tfrac{1}{2})\cdot\left(\frac{E_{\tau_0} - (\frac{1}{2} \pm \frac{1}{2})h\nu}{E_{\tau_0} + (\frac{1}{2} \mp \frac{1}{2})h\nu}\right)^v$$

where $v = 4\pi d/\lambda$.

The frequency ν_0 is related to the characteristic temperature Θ of the crystal by

$$h\nu_0 = k\Theta.$$

If the temperature of atom A is defined by $E_{\tau_0} = kt_A$, the preceding equation becomes

$$\gamma_{\tau_0}{}^\tau = (\pi^2/4lN^2)(m_A/m_B)(t_A/\Theta)(n_{\alpha\beta\gamma} + \tfrac{1}{2} \pm \tfrac{1}{2})\left(\frac{t_A - (1 \pm 1)\Theta l/N}{t_A + (1 \mp 1)\Theta l/N}\right)^V.$$

In forming the summation of $\gamma_{\tau_0}{}^\tau$ over all possible transitions, several more approximations will be made. First we replace the summation by the integration

$$3N^3\int_0^1 (2l/N)^2 d(2l/N).$$

Now if $\Theta \ll t_A$,

$$t_A/(t_A + \Theta 2l/N) \backsimeq (t_A - \Theta 2l/N)/t_A.$$

When the inequality is not satisfied, $n_{\alpha\beta\gamma}$ is very small except for those values of l for which the above is nearly valid, provided $t_B \leqq t_A$, which is usual in experiments. We thus have for the sum S of all the transition probabilities approximately

$$S = 3\pi^2(m_A/m_B)(t_A/\Theta)\int_0^1 (2l/N)(n_{\alpha\beta\gamma} + \tfrac{1}{2})(1 - 2l/N\Theta/t_A)^v d(2l/N).$$

If $\Theta > t_A$, the upper limit is to be replaced by t_A/Θ.

When $\Theta\ 2l/N \ll t_B$,

$$n_{\alpha\beta\gamma} + \tfrac{1}{2} = kt_B/h\nu_{\alpha\beta\gamma} = (t_B/\Theta)(N/2l).$$

However, since $(t_A - \Theta\ 2l/N)$ occurs to a fairly large power in the integrand, the major contributions to the integral are associated with small values of $2l/N$. Only a small error will thus be made if we substitute $(N/2l)t_B/\Theta$ for $n_{\alpha\beta\gamma} + 1/2$, provided t_B/Θ is not a very small fraction. We now have

$$S = 3\pi^2(m_A/m_B)(t_A/\theta)(t_B/\theta)\int_0^1 (1 - 2l/N\Theta/t_A)^{4\pi d/\lambda} d(2l/N). \qquad (8)$$

Integration gives

$$S = 3\pi^2(m_A/m_B)(t_A/\Theta)^2(t_B/\Theta)(1 + 4\pi d/\lambda)^{-1} \cdot [1 - (1 - \Theta/t_A)^{4\pi d/\lambda}]. \quad (9)$$

The second term in the last bracket is to be omitted if $\Theta/t_A > 1$.

We must investigate under what conditions we may substitute the S of (9) into (5). S has been obtained by a perturbation scheme which uses as an unperturbed solution that corresponding to an elastic reflection. Hence the method fails unless the unperturbed solution is the major portion of the final solution, which requires that $S \ll 1$. The previous analysis is thus limited to cases of large elastic reflection. However, a simplification is introduced by the large magnitude of the number of normal coordinates of the crystal. Consider a wave packet representing atom A being reflected by the crystal. We have seen that the majority of the transitions in the crystal are associated with an energy change very small in comparison with the energy of the wave packet, kt_A. This preferential weighting is caused by the presence of the factor $(1 - \Theta/t_A\ 2l/N)^{4\pi d/\lambda}$ in the integrand of (8). Moreover, the energy changes associated with various transitions partially cancel each other. Hence the character of the wave packet during collision is not very different from that corresponding to an elastic reflection, obtained by holding the lattice atoms rigidly to their positions of equilibrium.

The preceding discussion shows that in the collision of a wave packet with a crystal, the transition of one normal coordinate does not appreciably effect the probability that a second normal coordinate will suffer a change. Assuming that these transitions are absolutely independent, and defining β_k as the probability that the k'th normal coordinate suffers a transition, then the probability that all normal coordinates remain unchanged during collision is

$$U = \prod_k (1 - \beta_k).$$

Now the β_k's approach zero as N becomes infinite, so we may replace $(1 - \beta_k)$ by $\exp(-\beta_k)$. Thus

$$U = e^{-\Sigma \cdot \beta_k}.$$

To a good approximation the β_k of the wave packet will be equal to the probability calculated by the previous method that the k'th normal coordinate suffer a transition, irrespective of whether S is large or small. Since in that method only one normal coordinate can change during collision, β_k is simply the sum of the $|C_\tau|^2$'s for which the k'th quantum number changes by ± 1. The sum over the β_k's is thus equal to S. We have finally

$$U = e^{-S}. \quad (10)$$

This expression for U is no longer limited to small values of S, as is (5).

IV. Discussion

The one-dimensional collision of atoms with a crystal has been examined. An expression has been found for the fraction, U, of atoms reflected elasti-

cally. It is to be expected that the general dependence of U upon the parameters of the crystal and bombarding atoms will be approximately valid for actual reflections.

If the conditions $m_B \geqq 7m_A$, $\Theta \geqq 2t_A$, $\Theta \geqq 2t_B$, $d \backsimeq 0.25 \times 10^{-8}$ are satisfied, then $S \leqq 0.15$, and both (5) and (10) give $U \gtrsim 0.75$. Johnson[8] has found the experimental value of U to be as high as 0.5 for H atoms bombarding LiF. Here the above inequalities are valid. Application to other crystals gives consistently larger values for U than found by Johnson. It may be possible that the experimental values of U have been lowered by the presence of adsorbed gases on the crystal surface. The author wishes to acknowledge illuminating discussions with Dr. E. Wigner.

[8] T. H. Johnson, Phys. Rev. **35**, 650 (1930).

THE THEORY OF COMPLEX SPECTRA

By George H. Shortley
Palmer Physical Laboratory, Princeton University

(Received February 26, 1932)

Abstract

Part I. Calculation of the first order energies for configurations involving almost closed shells. A method is developed for calculating the electrostatic energy matrix for a configuration consisting of a closed shell minus ϵ electrons plus η other electrons with the same simplicity as for the corresponding configuration consisting of $\epsilon+\eta$ electrons. One classifies the states according to the quantum numbers of the "missing" and of the added electrons and calculates Slater's integral $(J-\delta K)$ for each pair of electrons in the usual way except that one attaches a negative sign to an integral connecting a "missing" with an added electron. It is then shown quite generally that the matrices of magnetic interaction are identical in LS coupling for the two configurations mentioned above except for a reversal of sign of the spin-orbit coupling parameter corresponding to the almost closed shell.

Part II. On two-electron spectra. It is shown that one is completely justified, when calculating the matrix of magnetic interaction in LS coupling by Johnson's method, in neglecting the equivalence of the two electrons of the atom. Similar results are shown to obtain for the addition of an electron to an ionic multiplet. The method of calculation of the matrix of magnetic interaction between configurations by this method is considered and it is shown that neglect of the equivalence of the electrons in case one of the configurations is composed of equivalent electrons leads to an error of a factor of $(2)^{1/2}$. One does not obtain a magnetic interaction between configurations which differ by more than the n value of one electron. This holds quite generally for any type of atom.

Part I. Calculation of the First Order Energies for Configurations Involving Almost Closed Shells

§1. Introduction

In 1929, Slater[1] gave a method of calculating the electrostatic interaction for any atom in terms of certain radial integrals (F's and G's). About a year ago Condon and the writer[2] extended Slater's tables to f electrons, and made a systematic comparison with this theory of all available data on two- and three-electron configurations. Johnson[3] has recently given a simple method of calculating the magnetic energy for any two-electron configuration, and he and others have made calculations of the magnetic energy for certain other configurations.[4] Finally, Condon and the writer[5] have made a

[1] J. C. Slater, Phys. Rev. **34**, 1293 (1929).

[2] E. U. Condon and G. H. Shortley, Phys. Rev. **37**, 1025 (1931).

[3] M. H. Johnson, Phys. Rev. **38**, 1628 (1931); see also part II of this paper.

[4] M. Johnson, Phys. Rev. **39**, 197 (1932); Goudsmit, Phys. Rev. **35**, 1325 (1930); Inglis, Phys. Rev. **38**, 862 (1931); M. H. Johnson, Phys. Rev. **40**, 127 (1932); J. H. Bartlett, Phys. Rev. **35**, 229 (1930).

[5] E. U. Condon and G. H. Shortley, Phys. Rev. **35**, 1342 (1930).

detailed comparison of experimental data with theory for *sp*, *sd*, *sf* configurations, using the calculations of electrostatic and magnetic energy as first given by Houston.[6]

These calculations are all for atoms at the left of the periodic table, *i.e.*, atoms whose configurations involve just a very few electrons outside of closed shells. In calculating the electrostatic energy by Slater's method one classifies the different possible complete sets[7] for a given configuration according to the values of M_L and M_S. The electrostatic energy is diagonal with respect to these z-components of angular momenta, so one proceeds to use the diagonal sum rule to find the energies for the different multiplets. The process of calculating the diagonal element of the electrostatic energy for a given complete set involves the evaluation, from tables, of a certain integral $(J-\delta K)$ for all possible pairs of individual sets in the complete set. This is not difficult when the configuration consists of just a few electrons, but the amount of work becomes tremendous when one deals with a configuration which involves an *almost closed shell*, particularly a *d* shell. For this reason, practically no calculations of electrostatic interaction for configurations of this type have been made.[8] Laporte and Inglis,[9] and simultaneously Condon and the writer[5] recognized on somewhat empirical grounds that Houston's formulas would apply to configurations of the type p^5s and d^9s if one inverted the level scheme and took X negative. The agreement with experiment was found to be remarkable. Heisenberg[10] has considered in general those configurations consisting purely of an almost closed shell, with no additional electrons, by considering the relation between the equations satisfied by n electrons and by n "Löchern" under the Pauli exclusion principle.

We shall here give a method by which the electrostatic energy for a configuration which consists of a closed shell minus ϵ electrons plus η other electrons can be calculated with the same simplicity as for the corresponding configuration consisting of ϵ electrons of the type of the closed shell plus the η other electrons, and shall show that the matrices of magnetic interaction (in LS coupling) are exactly the same for these two configurations except for a reversal of sign of the spin-orbit coupling parameter applying to the almost closed shell.

§2. Formulation of the problem

One takes as the zero order approximation for the calculation of atomic energies a model in which each electron moves independently in a central force field of potential energy $U(r)$; and considers as a perturbation the potential energy

[6] W. V. Houston, Phys. Rev. **33**, 297 (1929).

[7] A complete set of quantum numbers consists of N individual sets, each individual set specifying the nlm_lm_s of one of the N electrons of the configuration. These definitions are according to Condon, Phys. Rev. **36**, 1121 (1930).

[8] Slater gives partial results for p^5p; Ufford, in this laboratory, has calculated p^5p and d^9p.

[9] O. Laporte and D. R. Inglis, Phys. Rev. **35**, 1337 (1930).

[10] W. Heisenberg, Ann. d. Physik, **10**, 888, 1931.

$$V' = V - \sum_i U(r_i) = \sum_i \left[-\frac{Ze^2}{r_i} + \xi(r_i) \boldsymbol{L}_i \cdot \boldsymbol{S}_i - U(r_i) \right] + \sum_{i<j} \frac{e^2}{r_{ij}}, \quad (2.1)$$

where the first term represents the attraction between the electrons and the nucleus, the second the spin-orbit interaction, and the last the mutual repulsion of the electrons. In the central field $U(r)$, electron i with the individual set a^i of quantum numbers will have the eigenfunction[11]

$$u_i(a^i) = \frac{R_i(n^i l^i)}{r_i} \Theta_i(l^i m_l^i) \Phi_i(m_l^i) \delta(\sigma_i, m_s^i). \quad (2.2)$$

The zero order eigenfunction for the atom, going with the complete set $A = (a^1, a^2, \cdots, a^N)$ will be the antisymmetric combination

$$\psi(A) = \frac{1}{(N!)^{1/2}} \sum_P (-1)^p P u_1(a^1) u_2(a^2) \cdots u_N(a^N), \quad (2.3)$$

where P represents a permutation of quantum numbers relative to electron indices and p has the parity of P. (It is assumed that the Pauli exclusion principle is satisfied.) The zero order energy going with this ψ depends only on the *configuration*, i.e., on the set of $n^i l^i$ of A.

We are at the moment interested in the diagonal elements, in this representation, of the matrix of that part of V' which does not refer to the magnetic (spin-orbit) interaction. These terms fall into two classes: First we have

$$\sum_i \left[-\frac{Ze^2}{r_i} - U(r_i) \right], \quad (2.4)$$

which is of the form $\sum_i f(r_i)$. Slater has shown that for a given configuration the matrix of a quantity of this type is diagnoal, and that the value of the diagonal element depends only on the configuration; hence since at the present stage of the theory we are only interested in the disposition of the energy levels of a given configuration we need not consider these terms further. If it were of use, the calculation of the matrix of a quantity of this type is capable of a similar reduction to that which is made below for the quantity

$$G = \sum_{i<j} g(i, j) = \sum_{i<j} e^2/r_{ij}, \quad (2.5)$$

which is the second type of quantity occurring in V'.

The calculation of the relative electrostatic energies of the different multiplets in a configuration thus reduces to a calculation of the part of the matrix of G which refers to that configuration: in particular, it is necessary to calculate only the diagonal elements unless the configuration leads to more than one multiplet of a type. Slater has shown that the diagonal element of G for the state A is given by

[11] The different factors of this function, also the F's and G's which occur later, are as in Slater's paper, with the corrections suggested in reference 2, p. 1026.

$$A \mid G \mid A) = \sum_{k>t=1}^{N} \Big[\iint \bar{v}_1(a^k)\bar{v}_2(a^t)g(1,2)v_1(a^k)v_2(a^t)d\tau_1 d\tau_2$$

$$- \delta(m_s^k, m_s^t) \iint \bar{v}_1(a^k)\bar{v}_2(a^t)g(1,2)\, v_1(a^t)\, \bar{v}_2(a^k)\, d\tau_1 d\tau_2 \Big]$$

$$= \sum_{k>t} [J(k;t) - \delta(m_s^k m_s^t) K(k,t)], \tag{2.6}$$

where $v(a)$ is the spin-free eigenfunction $u(a)/\delta(\sigma, m_s)$. The first integral here is known as the direct integral, the second as the exchange integral. J and K are of course symmetric in k and t and independent of m_s^k and m_s^t.

§3. A summation over a closed shell

We shall need to consider a sum of terms like the summand of (2.6), in which a^t runs over the individual sets of the $n'l'$ closed shell, and in which a^k has the fixed value $nlm_l m_s$. We do not assume that all the sets a^t are different from a^k, as is the case in (2.6). This sum is, in detail

$$\sum_{m_s'} \sum_{m_l'=-l'}^{l'} \Big[\iint g(1,2) R_1^2(nl) R_2^2(n'l') \Theta_1^2(lm_l) \Theta_2^2(l'm_l')(2\pi)^{-2} d\tau_1 d\tau_2$$

$$- \delta(m_s, m_s') \iint g(1,2) R_1(nl) R_1(n'l') R_2(nl) R_2(n'l') \Theta_1(lm_l) \bar{\Phi}_1(m_l) \tag{3.1}$$

$$\cdot \Theta_1(l'm_l') \Phi_1(m_l') \Theta_2(lm_l) \Phi_2(m_l) \Theta_2(l'm_l') \bar{\Phi}_2(m_l') d\tau_1 d\tau_2 \Big].$$

Now the summation over m_s' gives a factor two in the first term and a factor one in the second. The summation over m_l' may be carried out by the use of the spherical harmonic addition theorem in the form

$$\sum_{m_l'=-l'}^{l'} \Theta_1(l'm_l') \Theta_2(l'm_l') \Phi_1(m_l') \bar{\Phi}_2(m_l') = \frac{2l'+1}{4\pi} P_{l'}(\cos\omega), \tag{3.2}$$

where ω is the angle between the radii vectores 1 and 2 from the origin; in particular if 1 and 2 are identical, $P_{l'}(\cos\omega) = 1$. This gives for our sum

$$2 \iint g(1,2) R_1^2(nl) R_2^2(n'l') \Theta_1^2(lm_l) \left(\frac{2l'+1}{4\pi}\right)\left(\frac{1}{2\pi}\right) d\tau_1 d\tau_2$$

$$- \iint g(1,2) R_1(nl) R_1(n'l') R_2(nl) R_2(n'l') \Theta_1(lm_l) \bar{\Phi}_1(m_l)$$

$$\cdot \Theta_2(lm_l) \Phi_2(m_l) \left(\frac{2l'+1}{4\pi}\right) P_{l'}(\cos\omega) d\tau_1 d\tau_2. \tag{3.3}$$

Now writing

$$g(1,2) = \frac{e^2}{r_{12}} = e^2 \sum_{\kappa=0}^{\infty} \frac{r_<^\kappa}{r_>^{\kappa+1}} P_\kappa(\cos\omega),$$

where $r_<$ is the lesser and $r_>$ the greater of r_1 and r_2, each of these integrals can be written as a sum over κ of integrals involving $e^2 r_<^{\kappa}/r_>^{\kappa+1} \cdot P_\kappa(\cos\omega)$ in place of $g(1, 2)$.

In the direct integral one may choose a new coordinate system for the variables θ_2, ϕ_2 with the direction θ_1, ϕ_1 as pole, in which case $\omega = \theta_2$. Then all terms vanish except for $\kappa = 0$, in which case the integration over θ_2, ϕ_2 gives a factor 4π. The integration over θ_1, ϕ_1 may then be performed directly, giving for the value of the direct integral

$$2(2l'+1)\int_0^\infty\int_0^\infty \frac{e^2}{r_>} R_1^2(nl)R_2^2(n'l')dr_1dr_2 = 2(2l'+1)F^0(nl;\, n'l') \quad (3.4)$$

or just the number of electrons in the closed shell times $F^0(nl;\, n'l')$.[11] This result is seen to be quite independent of m_l and m_s.

The exchange integral, on using the same expansion, becomes

$$-\sum_\kappa e^2 \iint \frac{r_<^\kappa}{r^{\kappa+1}} R_1(nl)R_1(n'l')R_2(nl)R_2(n'l')\Theta_1(lm_l)\bar{\Phi}_1(m_l)$$

$$\cdot\Theta_2(lm_l)\Phi_2(m_l)\left(\frac{2l'+1}{4\pi}\right)P_{l'}(\cos\omega)P_\kappa(\cos\omega)d\tau_1 d\tau_2$$

To evaluate this we expand the product $P_{l'}(\cos\omega)\, P_\kappa(\cos\omega)$ in a series of Legendre polynomials $P_\lambda(\cos\omega)$. This we may do since the $P_\lambda(\cos\omega)$ form a complete set of functions in the interval 0 to π. The coefficient of $P_\lambda(\cos\omega)$ in this expansion is given by

$$C^\lambda_{l'\kappa} = \frac{2\lambda+1}{2}\int_0^\pi P_\lambda(\cos\omega)P_{l'}(\cos\omega)P_\kappa(\cos\omega)\sin\omega d\omega$$

$$= \frac{2\lambda+1}{(2l'+1)^{1/2}(2\kappa+1)^{1/2}}\left\{b^\lambda(l'0;\,\kappa 0)\right\}^{1/2}$$

where the last factor is the positive square root of b^λ as given in the tables of Slater and of Condon and Shortley.[12]

When we substitute the relation

$$P_{l'}(\cos\omega)P_\kappa(\cos\omega) = \sum_\lambda C^\lambda_{l'\kappa}P_\lambda(\cos\omega),$$

the above exchange integral becomes a double sum over κ and λ. In each summand the dependence on θ_2, ϕ_2 is through the factor

$$\Theta_2(lm_l)\Phi_2(m_l)P_\lambda(\cos\omega).$$

The integral of this over the whole range of θ_2 and ϕ_2 vanishes unless $\lambda = l$, in which case it becomes[13]

[12] A closed expression for the integral of the product of three ordinary Legendre polynomials in known. See Hobson, *Spherical Harmonics*, p. 87, or Gaunt, Phil. Trans. Roy. Soc. **A228**, 195 (1929).

[13] This is a well known integral in the theory of spherical harmonics, see, *e.g.*, Mac Robert, *Spherical Harmonics*, p. 137.

$$\frac{4\pi}{2l+1}\Theta_1(lm_l)\Phi_1(m_l).$$

The integration over θ_1, ϕ_1 can now be performed to give simply a constant factor. Thus we are left with

$$-\frac{2l'+1}{2l+1}\sum_\kappa C^l_{l'\kappa 0^2}\int_0^\infty\int_0^\infty\frac{r_<^\kappa}{r_>^{\kappa+1}}R_1(nl)R_1(n'l')R_2(nl)R_2(n'l')dr_1dr_2$$

$$=-\frac{2l'+1}{2l+1}\sum_\kappa C^l_{l'\kappa}G^\kappa(nl;\,n'l')$$

for the value of the exchange integral. Note that this result is also independent of m_l and m_s. The values of κ which can occur in the summation are those which satisfy the triangular conditions

$$\kappa+l+l'=2g\quad(g\text{ integral})$$
$$|l-l'|\leqq\kappa\leqq l+l'.$$

This calculation, incidentally, furnishes a formal direct proof of the fact that the electrostatic interaction of any electron with a closed shell is independent of the m_l, m_s of that electron.

§4. Calculation of the matrix of electrostatic energy

Consider the state A of a configuration which consists of a closed shell minus ϵ electrons plus η other electrons in addition to any number of completely closed shells. Denote the individual sets of the closed shell by a^1, a^2, $\cdots$, a^ϵ, $a^{\epsilon+1}$, $\cdots$, a^n, where $a^{\epsilon+1}$, $\cdots$, a^n are the sets occurring in A and $a^1\cdots a^\epsilon$ are the "missing" sets. Denote the rest of the sets occurring in A by b^1, b^2, $\cdots$, b^η. In the diagonal element (2.6) of G we will have first a sum over all pairs of electrons belonging to the closed shells present plus a sum of terms between each of the other electrons and all electrons in closed shells. These are just as in Slater's calculations—dependent only on the configuration—and may be given in terms of simple expressions involving radial integrals. In particular we have found in the last section the expression for a sum of the latter type. The remainder of the diagonal element is given by

$$\left.\begin{aligned}&\sum_{i>j=\epsilon+1}^{n}\left[J(a^ia^j)-\delta(m_s^i,m_s^j)K(a^ia^j)\right]&(\alpha)\\+&\sum_{i=\epsilon+1}^{n}\sum_{j=1}^{\eta}\left[J(a^ib^j)-\delta(m_s^i,m_s^j)K(a^ib^j)\right]&(\beta)\\+&\sum_{i>j=1}^{\eta}\left[J(b^ib^j)-\delta(m_s^i,m_s^j)K(b^ib^j)\right].&(\gamma)\end{aligned}\right\}\quad(4.1)$$

Call these three terms (α), (β), (γ) respectively. We wish now to reduce the inconvenient summations over the range $\epsilon+1$ to n to the more convenient summations over the range 1 to ϵ.

Consider first the term (α). We may write

$$
\begin{aligned}
(\alpha) &= \sum_{i>j=\epsilon+1}^{n} [J(a^i a^j) - \delta(m_s^i, m_s^j) K(a^i a^j)] \\
&= \sum_{i>j=1}^{n} [\cdots] - \sum_{i=\epsilon+1}^{n} \sum_{j=1}^{\epsilon} [\cdots] - \sum_{i>j=1}^{\epsilon} [\cdots] \\
&= \sum_{i>j=1}^{n} [\cdots] - \sum_{i=1}^{n} \sum_{j=1}^{\epsilon} [\cdots] + \sum_{i=1}^{\epsilon} \sum_{j=1}^{\epsilon} [\cdots] - \sum_{i>j=1}^{\epsilon} [\cdots] \\
&= \sum_{i>j=1}^{n} [J(a^i a^j) - \delta(m_s^i, m_s^j) K(a^i a^j)] \\
&\quad - \sum_{i=1}^{n} \sum_{j=1}^{\epsilon} [J(a^i a^j) - \delta(m_s^i, m_s^j) K(a^i a^j)] \qquad (4.2) \\
&\quad + \sum_{i>j=1}^{\epsilon} [J(a^i a^j) - \delta(m_s^i, m_s^j) K(a^i a^j)] + \sum_{i=1}^{\epsilon} [J(a^i a^i) - K(a^i a^i)].
\end{aligned}
$$

Now since $J(a^i a^i) \equiv K(a^i a^i)$ the last term vanishes identically. The first term is a summation over the closed shell and so depends only on the configuration. The second term is of the form of the sum of terms like (3.1) in which $nlm_l m_s$ takes on the values 1 to ϵ. Since (3.1) was shown to be independent of m_l and m_s, this term depends only on the configuration. Hence the only term of (α) which may vary within the configuration is

$$
\sum_{i>j=1}^{\epsilon} [J(a^i a^j) - \delta(m_s^i, m_s^j) K(a^i a^j)], \qquad (4.3)
$$

which has the required simple summation.

The term (β) becomes similarly

$$
\begin{aligned}
(\beta) &= \sum_{i=\epsilon+1}^{n} \sum_{j=1}^{\eta} [J(a^i b^j) - \delta(m_s^i, m_s^j) K(a^i b^j)] \\
&= \sum_{i=1}^{n} \sum_{j=1}^{\eta} [J(a^i b^j) - \delta(m_s^i, m_s^j) K(a^i b^j)] \qquad (4.4) \\
&\quad - \sum_{i=1}^{\epsilon} \sum_{j=1}^{\eta} [J(a^i b^j) - \delta(m_s^i, m_s^j) K(a^i b^j)]
\end{aligned}
$$

of which the first term depends only on the configuration.

Since the term (γ) is already in its simplest form we find for *that part of* $(A\,|G\,|A)$ *which may vary within the configuration*, the value

$$
\left.
\begin{aligned}
&\sum_{i>j=1}^{\epsilon} [J(a^i a^j) - \delta(m_s^i, m_s^j) K(a^i a^j)] && (\alpha) \\
&- \sum_{i=1}^{\epsilon} \sum_{j=1}^{\eta} [J(a^i b^j) - \delta(m_s^i, m_s^j) K(a^i b^j)] && (\beta) \\
&+ \sum_{i>j=1}^{\eta} [J(b^i b^j) - \delta(m_s^i, m_s^j) K(b^i b^j)]. && (\gamma)
\end{aligned}
\right\} \qquad (4.5)
$$

This gives a perfectly simple and definite procedure for calculating the electrostatic energies for such a configuration. Since $\sum m_l$ and $\sum m_s$ over a closed shell is zero one can proceed as follows:

Calculation of the electrostatic energy for a configuration consisting of a closed shell minus ϵ *electrons plus* η *other electrons* (call the group of ϵ missing electrons group a, the group of η other electrons group b):

I. Make a table of all allowed complete sets of quantum numbers for the configuration consisting of groups a and b. Call this Table I, and label the different complete sets A, B, C, $\cdots$. This table lists all possible complete sets for the configuration in question, each labelled by the quantum numbers of the "missing" electrons and the additional electrons.

II. Make a second table which is identical with the first except that the signs of m_l and m_s for the electrons of group a are reversed. Denote the entries which correspond to A, B, $C \cdots$ by A_0, B_0, $C_0 \cdots$. The values of $\sum m_l$ and $\sum m_s$ as obtained from *this table* will be the true values for the states A, B, $C \cdots$ of the configuration we are calculating. Call this Table II.

III. Combine the values of M_L and M_S *from Table II* into a double entry table as in Slater, determine the multiplets, and arrange for their calculation by the diagonal sum method.

IV. Calculate the $\sum [J - \delta K]$ for each complete set as outlined in Slater, *using the quantum numbers of Table I, and assigning a plus sign if both electrons lie in group* a *or group* b, *but a minus sign if one electron lies in each group*, in accordance with (4.5).

This procedure involves exactly the same amount of calculation as for the corresponding configuration with ϵ electrons *toward* a closed shell.[14] However, because of the various reversals of sign that occur the results will in general be quite different in appearance from the corresponding results for the analogous simple configurations.

As an illustration we will sketch the calculation of the electrostatic energy matrix for p^5p. The missing electron a then belongs to the shell n^ap, the added electron b to the shell n^bp. There are nine possible assignments of m_l values and going with each of these there are four possible assignments of m_s values. These are listed below along with the *true values* of M_L and M_S.

m_l^a	m_l^b	M_L	m_s^a	m_s^b	M_S
− 1	1	2 (α)	−	+	1 (1)
	0	1 (β)	−	−	0 (2)
	− 1	0 (γ)	+	+	0 (3)
0	1	1 (δ)	+	−	− 1 (4)
	0	0 (ζ)			
	− 1	− 1 (θ)			
1	1	0 (ι)			
	0	− 1 (κ)			
	− 1	− 2 (λ)			

[14] Table II is not actually constructed: one simply calculates the true M_L and M_S from Table I by reversing the signs of the electrons of group a. Table II will be useful in the considerations of §5.

For the calculation of the electrostatic interaction sets with M_L, $M_S \geqq 0$ are arranged in the following double-entry table.

M_L	$M_S = 0$	$M_S = 1$
2	$\alpha_2\alpha_3$	α_1
1	$\beta_2\beta_3$	β_1
	$\delta_2\delta_3$	δ_1
0	$\gamma_2\gamma_3$	γ_1
	$\zeta_2\zeta_3$	ζ_1
	$\iota_2\iota_3$	ι_1

Using Slater's table of a's and b's and the signs as directed above, we obtain the electrostatic energies[15] as follows

$$
\begin{aligned}
{}^3D = (\alpha_1) &= -F_0 - F_2 && {}^3D = -F_0 - F_2\\
{}^1D + {}^3D = (\alpha_2) + (\alpha_3) &= -F_0 - F_2 + 6G_2 && {}^1D = -F_0 - F_2 + 12G_2\\
{}^3D + {}^3P = (\beta_1) + (\delta_1) &= -2F_0 + 4F_2 && {}^3P = -F_0 + 5F_2\\
{}^3P + {}^3D + {}^1D + {}^1P = (\beta_2) + (\beta_3) + (\delta_2) + (\delta_3)&\\
&= -4F_0 + 8F_2 + 12G_2 && {}^1P = --F_0 + 5F_2\\
{}^3S + {}^3P + {}^3D = (\gamma_1) + (\zeta_1) + (\iota_1)&\\
&= -3F_0 - 6F_2 && {}^3S = -F_0 - 10F_2\\
{}^3S + {}^3P + {}^3D + {}^1S + {}^1P + {}^1D = (\gamma_2) + (\gamma_3) + (\zeta_2) + (\zeta_3) + (\iota_2) + (\iota_3)&\\
&= -6F_0 - 12F_2 + 6G_0 + 12G_2 && {}^1S = -F_0 - 10F_2 + 6G_0
\end{aligned}
$$

where

$$
\begin{aligned}
F_0 &= F^0(n^a p;\, n^b p)\\
F_2 &= (1/25)F^2(n^a p;\, n^b p)\\
G_0 &= G^0(n^a p;\, n^b p)\\
G_2 &= (1/25)G^2(n^a p;\, n^b p).
\end{aligned}
$$

These energies are seen to be quite different from those for the corresponding configuration pp as on p. 1029, reference 2, especially in regard to the singlet-triplet separations.

One can generalize this scheme to include configurations which involve two or more almost closed shells.* The procedure is essentially the same as above with the additional direction that one assign a plus sign to the integral $J - \delta K$ connecting electrons "missing" from two different shells.

[15] These energies check with those calculated by Ufford (unpublished) for the same configuration by the straightforward method. The F's and G's with subscripts are defined as on p. 1026, reference 2, to be Slater's F's and G's divided by the denominator of a^k or b^k as it occurs in the tables.

* This case is not completely uninteresting from the spectroscopic point of view: it furnishes the most pleasing way of inferring that the electrostatic energies for the configurations p^5s and d^9s are given by the same formulas as for the configurations ps and ds.

§5. The matrix of magnetic energy

In this section we shall show quite generally that the matrix of the spin orbit interaction

$$V^1 = \sum_i V_i^1 = \sum_i \xi(r_i) \boldsymbol{L}_i \cdot \boldsymbol{S}_i \tag{5.1}$$

is the same, in LS coupling,[16] for the configuration consisting of a closed $n^a l^a$ shell *minus* ϵ electrons plus η other electrons as for the configuration consisting of ϵ electrons $n^a l^a$ plus the η other electrons, except for a reversal of the sign of the spin orbit coupling parameter

$$\zeta_{n^a l^a} = (n^a l^a \mid \xi(r) \mid n^a l^a). \tag{5.2}$$

For convenience we shall speak of these two configurations as "configuration $\mathcal{R}$" and "configuration $\mathcal{L}$" (right and left in the periodic table). Our procedure will be as follows: To each of the zero order eigenfunctions of configuration $\mathcal{R}$ with a definitely chosen phase, we correlate a zero order eigenfunction of configuration $\mathcal{L}$ in a definite way. We shall then see that the matrix of V^1 in these two configurations differs only in the sign of $\zeta_{n^a l^a}$, while the matrices of $\boldsymbol{L}^2$, $\boldsymbol{S}^2$, and $\boldsymbol{L} \cdot \boldsymbol{S}$ are identical. The latter fact shows that the transformation to LS coupling will be the same for the two configurations; and the former, that the matrices of V^1, when so transformed, will differ only in the sign of $\zeta_{n^a l^a}$. Now the matrices of magnetic interaction have been calculated[4] in LS coupling for various configurations of the type $\mathcal{L}$; in particular for all two electron configurations,[17] which correspond to the important cases of the type p^5x, d^9x, etc. Hence this calculation, which is in general rather complicated, will not have to be made independently for configurations of the type $\mathcal{R}$.

According to this program, we will first adopt a standard order of listing quantum numbers in the eigenfunction (2.3). The individual sets will be listed first in increasing order of the n values; the individual sets with a particular n will be arranged in increasing order of the l values; those with particular n and l will be listed in decreasing order of the m_l values; and the set with $m_s = +\frac{1}{2}$ will be listed before that with $m_s = -\frac{1}{2}$ in case the two sets agree in regard to n, l, and m_l. With the individual sets arranged in this order in (2.3) we obtain what we shall call $\psi(A)$. We shall now wish to choose the phase of the zero order eigenfunctions of configuration $\mathcal{R}$ in a particular way. For the complete set A we shall take $+\psi(A)$ if the sum of the m_l values of the electrons in the almost completed shell is even, $-\psi(A)$ if this sum is odd. We shall correlate to each complete set A of configuration $\mathcal{R}$ in Table I the complete set A_0 of configuration $\mathcal{L}$ as given in Table II, the eigenfunction

[16] The fact that we determine this matrix in LS coupling does not mean that the energy calculation is in any way limited to the Russell-Saunders case. The energy matrix will not be *diagonal* in LS coupling, and the transformation necessary to diagonalize it will represent the socalled *breakdown* of the coupling.

[17] M. H. Johnson, reference 3; this method needs a justification, however, which is given in Part II of this paper.

being taken with the + phase when the set is in the standard order. These two complete sets have the same M_L and M_S values.

We must now verify our assertions regarding the matrices of V^1, $\boldsymbol{L}^2$, $\boldsymbol{S}^2$, and $\boldsymbol{L}\cdot\boldsymbol{S}$. Formulas for the matrix elements of such quantities have been obtained by Johnson[4] using the general formulas given by Condon.[7] These formulas will have to be supplemented by an exact determination of the sign of the non-diagonal elements. A general rule may be stated as follows.[18] To the matrix component between two states, $\psi(A)$ and $\psi(A')$, differing in regard to one or two individual sets, one prefixes + or − according to the parity of the permutation which changes the set A' from its standard order to the order in which equal elements in A and A' occupy the same places.

We will use the following notation for the matrix components of angular momentum for one electron.

$$\begin{aligned}(c\,|\,L_x\,|\,d) &= (n^c l^c m_l^c m_s^c\,|\,L_x\,|\,n^d l^d m_l^d m_s^d)\\ &= \delta(n^c l^c m_s^c;\, n^d l^d m_s^d)\delta(m_l^c \pm 1, m_l^d)\tfrac{1}{2}\hbar[(l_d \pm m_l^d)(l^d \mp m_l^d + 1)]^{1/2};\end{aligned} \tag{5.3}$$

$$(c\,|\,S_x\,|\,d) = \delta(n^c l^c m_l^c;\, n^d l^d m_l^d)\delta(m_s^c \pm 1, m_s^d)\tfrac{1}{2}\hbar; \tag{5.4}$$

$$\begin{aligned}(c\,|\,\boldsymbol{L}\cdot\boldsymbol{S}\,|\,d) &= \hbar^2\delta(n^c l^c m_j^c;\, n^d l^d m_j^d)\{\delta(c,d)m_l^c m_s^c + \tfrac{1}{2}\delta(m_s^c, m_s^d \pm 1)\\ &\quad\cdot[(l_c - m_j^c + \tfrac{1}{2})(l_c + m_j^c + \tfrac{1}{2})]^{1/2}\};\end{aligned} \tag{5.5}$$

where $m_j = m_l + m_s$. One fact we shall need particularly is that the matrix component of any of these quantities between the states $m_l^c m_s^c$ and $m_l^d m_s^d$ is the same as that between $-m_l^c$, $-m_s^c$ and $-m_l^d$, $-m_s^d$.

Matrix of V^1. One obtains a non-diagonal element of V^1 only between states differing in regard to one individual set, say that c occurs in A while c' occurs in A'. The value of this component is

$$\pm\,\zeta_{n^c l^c}(c\,|\,\boldsymbol{L}\cdot\boldsymbol{S}\,|\,c'), \tag{5.6}$$

where the sign is determined in accordance with the above convention. If the element c occurs in the group b of additional electrons one obtains the same value for configuration $\mathcal{R}$ and configuration $\mathcal{L}$. If c is in the almost closed shell c and c' must be of the form $(m_l)^-$ and $(m_l-1)^+$ respectively. A and A' are already in the same order, as are A_0 and A_0'. But $\psi(A)$ and $\psi(A')$ are taken with opposite signs by our rule for choosing the phase. Hence in the matrix components

$$(A\,|\,V^1\,|\,A') = -\,\zeta_{n^a l^a}(m_l, -\,|\,\boldsymbol{L}\cdot\boldsymbol{S}\,|\,m_l - 1, +),$$

$$(A_0\,|\,V^1\,|\,A_0') = \zeta_{n^a l^a}(-\,m_l + 1, -\,|\,\boldsymbol{L}\cdot\boldsymbol{S}\,|\,-\,m_l, +) = -\,(A\,|\,V^1\,|\,A'),$$

$\zeta_{n^a l^a}$ occurs with opposite sign. For the diagonal element of V^1 one has

[18] Communicated by Professor Condon.

$$(A \mid V^1 \mid A) = \zeta_{n^a l^a} \sum_{i=\epsilon+1}^{n} m_l^i m_s^i + \sum_{i=1}^{\eta} (b^i \mid V_i^1 \mid b^i)$$

$$= \zeta_{n^a l^a} \sum_{i=1}^{n} m_l^i m_s^i - \zeta_{n^a l^a} \sum_{i=1}^{\epsilon} m_l^i m_s^i + \sum_{i=1}^{\eta} (b^i \mid V_i^1 \mid b^i),$$

where the first sum is zero. For the configuration $\mathcal{L}$ one obtains

$$(A_0 \mid V^1 \mid A_0) = \zeta_{n^a l^a} \sum_{i=1}^{\epsilon} m_l^i m_s^i + \sum_{i=1}^{\eta} (b^i \mid V_i^1 \mid b^i)$$

since the product $m_l^i m_s^i$ is the same for a missing electron as for the corresponding electron in A_0. These two elements are seen to be the same except for the sign of $\zeta_{n^a l^a}$. This completes the proof of our assertion concerning the matrices of V^1.

Matrix of $\boldsymbol{L}^2$. We get non-diagonal matrix components of $\boldsymbol{L}^2$ only between states differing in regard to two individual sets, say c and d. This component has the value

$$(A \mid \boldsymbol{L}^2 \mid A') = \pm 4\{(c \mid L_x \mid c')(d \mid L_x \mid d') - (c \mid L_x \mid d')(d \mid L_x \mid c')\}.$$

Now if both c and c' are in group b the value is obviously the same for $\mathcal{R}$ and $\mathcal{L}$. If c is in shell a, d in group b, one obtains no value from the second term. In order to obtain a value for the first term c and c' must be of the form $(m_l)^{\pm}$, $(m_l - 1)^{\mp}$. If the element lying between these two is present in $\mathcal{R}$, so far as the shell a is concerned A and A' differ by an odd permutation, while A_0 and A_0' do not differ in order. The converse is true if this element is missing. This introduces one difference of sign between $(A \mid \boldsymbol{L}^2 \mid A')$ and $(A_0 \mid \boldsymbol{L}^2 \mid A_0')$; however we get a second difference of sign from the different choice of phase of $\psi(A)$ and $\psi(A')$, which makes these two elements just equal. In a similar way one demonstrates the equality in case c and c' are both in shell a.

The diagonal element of $\boldsymbol{L}^2$ is

$$(A \mid \boldsymbol{L}^2 \mid A) = \hbar^2 M_L^2 + \hbar \sum_{c} \{l^c(l^c + 1) - (m_l^c)^2\} - 4 \sum_{c<d} (c \mid L_x \mid d)^2,$$

where c and d run over all the individual sets of A. The first term is the same for A and A_0, as are also the parts of the second and third terms arising from group b. The part of these terms which arises from group a is the same if calculated for A_0 as for the group of *missing* electrons. The equivalence of this to the calculation for $\mathcal{L}$ is a direct consequence of the following interesting relation: If one takes the integers $l, \cdots, 0, \cdots, -l$ and arranges them into two groups α and β, then the following sum has the same value for groups α and β (the individual integers are denoted by m_l):

$$\sum_{m_l} [l(l+1) - m_l^2] + \sum [-l(l+1) + m_l^2 - m_l],$$

where the second sum is taken over those m_l's for which $m_l - 1$ is also in the group.

This completes the proof that the matrix of $\boldsymbol{L}^2$ is the same for $\mathcal{R}$ and $\mathcal{L}$. The calculation for $\boldsymbol{S}^2$ is very similar and will not be discussed in detail.

Matrix of $\boldsymbol{L}\cdot\boldsymbol{S}$. If A' differs from A in two electrons, the calculation is much the same as for $\boldsymbol{L}^2$ and $\boldsymbol{S}^2$. If A differs from A' in regard to one electron, c, one obtains[19]

$$(A\,|\,\boldsymbol{L}\cdot\boldsymbol{S}\,|\,A') = \pm\{(c\,|\,\boldsymbol{L}\cdot\boldsymbol{S}\,|\,c') - 2\sum_d (c\,|\,L_x\,|\,d)(d\,|\,S_x\,|\,c') - 2\sum_{d'}(c'\,|\,L_x\,|\,d')(d'\,|\,S_x\,|\,c)\},$$

where d and d' run over all the individual sets common to A and A'. If c is in group b we obviously get the same result for $\mathcal{L}$ and $\mathcal{R}$. If c is in shell a, we get no value unless c, c' are of the form $(m_l)^-$, $(m_l-1)^+$. Then to obtain a value for the second term d must be $(m_l-1)^-$, for the third term d' must be $(m_l)^+$. With these values for d and d', the second and third terms each become $-(c\,|\boldsymbol{L}\cdot\boldsymbol{S}\,|c')$. One may easily verify that A and A', A_0 and A_0' always differ by the same permutation. One obtains a minus sign in the element $(A\,|\boldsymbol{L}\cdot\boldsymbol{S}|A')$ due to the difference in phase in the eigenfunctions. Then if neither d nor d' is in configuration $\mathcal{L}$, they must both be in $\mathcal{R}$, hence in this case

$$(A\,|\,\boldsymbol{L}\cdot\boldsymbol{S}\,|\,A') = -(c\,|\,\boldsymbol{L}\cdot\boldsymbol{S}\,|\,c')$$
$$(A_0\,|\,\boldsymbol{L}\cdot\boldsymbol{S}\,|\,A_0') = (c\,|\,\boldsymbol{L}\cdot\boldsymbol{S}\,|\,c') - 2(c\,|\,\boldsymbol{L}\cdot\boldsymbol{S}\,|\,c') = (A\,|\,\boldsymbol{L}\cdot\boldsymbol{S}\,|\,A').$$

Similar considerations hold for the other two possibilities.

The diagonal element of $\boldsymbol{L}\cdot\boldsymbol{S}$ is given by

$$(A\,|\,\boldsymbol{L}\cdot\boldsymbol{S}\,|\,A') = M_L M_S \hbar^2,$$

which obviously equals $(A_0\,|\boldsymbol{L}\cdot\boldsymbol{S}\,|A_0')$. This completes the proof as outlined at the beginning of the section.

We will not at this time go further into detail concerning calculations of special cases for configurations of this type. The writer is preparing a sequel to this paper which will consider all the special cases of interest and their comparison with experiment.

PART II. ON TWO-ELECTRON SPECTRA

§6. The direct calculation of the magnetic interaction in *LS* coupling

In a recent paper M. H. Johnson[3] gave a clever method for the calculation of the matrix of magnetic interaction directly in *LS* coupling for any two-electron configuration. This method, which was based on the work of Güttinger and Pauli[20] on the calculation of angular momentum matrices, at once disposed of the whole of the two-electron problem which previously had been attacked configuration by configuration, by Houston,[6] Bartlett,[4] and Goudsmit.[4] However, in making this calculation, Johnson assumes without

[19] This formula is alternative to, and somewhat simpler than, that given on p. 199 of Johnson's paper, reference 4.

[20] P. Güttinger and W. Pauli, Zeits. f. Physik **67**, 754 (1931).

proof[21] that one will obtain correct results by calculating the matrix of $\xi(r_1)\boldsymbol{L}_1\cdot\boldsymbol{S}_1+\xi(r_2)\boldsymbol{L}_2\cdot\boldsymbol{S}_2$ in the system of states characterized by definite values nls of the quantum numbers of electron 1, definite values $n'l's'$ of the quantum numbers of electron 2, and the values $LSJM_J$ of the resultant momenta; of course calculating nothing at all for states known to be ruled out by the exclusion principle. This complete neglect of the equivalence of the two electrons, which is usually taken care of by antisymmetrization, needs justification; and since our calculations of magnetic interaction for configurations of the type p^5x, d^9x, etc. depend on the use of Johnson's matrices, it was thought appropriate to include a justification of this point at this place.

We are interested in the matrix components of $\xi(r_1)\boldsymbol{L}_1\cdot\boldsymbol{S}_1+\xi(r_2)\boldsymbol{L}_2\cdot\boldsymbol{S}_2$ joining antisymmetric states characterized by $LSJM_J$ for a given two-electron configuration. Each such state is a definite known[22] linear combination of states characterized by LSM_LM_S, or say γ for short. Now our basic states are characterized by $n_1l_1m_{l1}s_1m_{s1}$; $n_2'l_2'm_{l2}'s_2{}^4m_{s2}'$. Call the eigenfunction going with this state $\phi(n_1\cdots m_{s2}')$. We can build up a state characterized by $n_1l_1s_1$; $n_2'l_2's_2'$; LSM_LM_S by taking definite known linear combinations of these states ϕ. Call the state so obtained $\psi(\alpha_1,\beta_2,\gamma)$. Now this state will be neither symmetric nor antisymmetric in 1 and 2 if $\beta\neq\alpha$, since if it were it would be an eigenstate for both quantum numbers α and quantum numbers β for each electron, and this is of course impossible. However, to have a physical significance we must make this state antisymmetric in 1 and 2. Now since the state $\psi(\alpha_2,\beta_1,\gamma)$ is also an eigenstate of γ, the antisymmetric combination

$$\Psi(\gamma)=\frac{1}{(2)^{1/2}}[\psi(\alpha_1,\beta_2,\gamma)-\psi(\alpha_2,\beta_1,\gamma)] \tag{6.1}$$

will be an eigenstate of γ and will be the only one that satisfies the antisymmetry condition. We can then consider the state $LSJM_J$ in which we are interested as built up of linear combinations, in the known way, of these antisymmetric states. The general matrix component joining two states $LSJM_J$; $L'S'J'M_J'$ of the same configuration will be the sum of matrix components between these antisymmetric functions, with a constant α and β. The most general term will be

$$\tfrac{1}{2}[\bar\psi(\alpha_1,\beta_2,\gamma)-\bar\psi(\alpha_2,\beta_1,\gamma)][\{1\}+\{2\}][\psi(\alpha_1,\beta_2,\gamma')-\psi(\alpha_2,\beta_1,\gamma')]$$

where $\{1\}=\xi(r_1)\boldsymbol{L}_1\cdot\boldsymbol{S}_1$, etc. Expanded, this becomes

$$\begin{aligned}
&\tfrac{1}{2}\bar\psi(\alpha_1,\beta_2,\gamma)\{1\}\psi(\alpha_1,\beta_2,\gamma')+\tfrac{1}{2}\bar\psi(\alpha_2,\beta_1,\gamma)\{1\}\psi(\alpha_2,\beta_1,\gamma')\\
&+\tfrac{1}{2}\bar\psi(\alpha_1,\beta_2,\gamma)\{2\}\psi(\alpha_1,\beta_2,\gamma')+\tfrac{1}{2}\bar\psi(\alpha_2,\beta_1,\gamma)\{2\}\psi(\alpha_2,\beta_1,\gamma')\\
&-\tfrac{1}{2}\bar\psi(\alpha_1,\beta_2,\gamma)\{1\}\psi(\alpha_2,\beta_1,\gamma')-\tfrac{1}{2}\bar\psi(\alpha_2,\beta_1,\gamma)\{1\}\psi(\alpha_1,\beta_2,\gamma')\\
&-\tfrac{1}{2}\bar\psi(\alpha_1,\beta_2,\gamma)\{2\}\psi(\alpha_2,\beta_1,\gamma')-\tfrac{1}{2}\bar\psi(\alpha_2,\beta_1,\gamma)\{2\}\psi(\alpha_1,\beta_2,\gamma').
\end{aligned} \tag{6.2}$$

[21] I am indebted to Mr. Johnson for calling my attention to this point.

[22] The matrix of this transformation was first given in general by E. P. Wigner, *Gruppentheorie*, p. 206.

Now of this the first term equals the fourth and the second the third, since the interchange of the electron indices 1 and 2 in a matrix component does not affect its value. Together these first four terms give

$$\bar{\psi}(\alpha_1, \beta_2, \gamma)[\{1\} + \{2\}]\psi(\alpha_1, \beta_2, \gamma'). \tag{6.3}$$

Now as to the last four terms; certainly none of these will have any value unless $\beta \equiv \alpha$, *i.e., unless the electrons are equivalent,* since any function of electron 1 is diagonal with respect to the quantum numbers of electron 2 and *vice versa*. If the electrons are equivalent, $\psi(\alpha_1, \alpha_2, \gamma)$ will be either symmetric or antisymmetric. To show this we note that $\psi(\alpha_1, \alpha_2, \gamma)$ is the product of an orbital function[23] $\psi_r(n_1l_1, n_2l_2, LM_L)$ and a spin function $\psi_\sigma(s_1, s_2, SM_S)$. We express this ψ_r in terms of the fundamental ϕ_r's by the transformation

$$\psi_r(n_1l_1n_2l_2LM_L) = \sum_{m_l, m_l'} \phi_r(n_1l_1m_{l1}n_2l_2m'_{l2})(n_1l_1m_{l1}n_2l_2m'_{l2} \mid n_1l_1n_2l_2LM_L). \tag{6.4}$$

Now from page 206 of Wigner we have the relation

$$(n_1l_1m'_{l1}n_1l_1m_{l2} \mid n_1l_1n_2l_2LM_L) = (-1)^{2l-L}(n_1l_1m_{l1}n_2l_2m'_{l2} \mid n_1l_1n_2l_2LM_L). \tag{6.5}$$

In the summation (6.4) m_l and m_l' run over the same range $-l \cdots l$. Hence we may rewrite our summation, using (6.5), as follows:

$$\begin{aligned}\psi_r(n_1l_1n_2l_2LM_L) = \sum_{m_l < m_l'} &[\phi_r(n_1l_1m_{l1}n_2l_2m'_{l2}) + (-1)^{2l-L}\phi_r(n_1l_1m'_{l1}n_2l_2m_{l2})] \\ &\cdot (n_1l_1m_{l1}n_2l_2m'_{l2} \mid n_1l_1n_2l_2LM_L) \\ &+ \sum_{m_l} \phi_r(n_1l_1m_{l1}n_2l_2m_{l2})(n_1l_1m_{l1}n_2l_2m_{l2} \mid n_1l_1n_2l_2LM_L),\end{aligned}$$

and this, since ϕ_r is merely a product of a function of electron 1 with one of electron 2, is symmetric if $2l-L$ is even and antisymmetric if $2l-L$ is odd. Similarly ψ_σ is symmetric if $2s-S$ is even, antisymmetric if it is odd. Hence $\psi(\alpha_1, \alpha_2, \gamma)$, the product of ψ_r and ψ_σ, will be either symmetric or antisymmetric depending on whether $2(l+s)-(L+S)$ is even or odd.

If either $\psi(\alpha_1, \alpha_2, \gamma)$ or $\psi(\alpha_1, \alpha_2, \gamma')$ is symmetric we get zero for our matrix component. (Either $\Psi(\gamma)$ or $\Psi(\gamma')$ vanishes identically; the last four terms of (6.2) cancel the first four.) If both are antisymmetric the last four terms of (6.2) are exactly equal to the first four. But in this case our normalization of Ψ is wrong by a factor $1/(2)^{1/2}$ so again we get just the result (6.3).

Hence, for any case in which states γ and γ' are permitted by the exclusion principle we have the result

$$\bar{\Psi}(\gamma)[\{1\} + \{2\}]\Psi(\gamma') = \bar{\psi}(\alpha_1, \beta_2, \gamma)[\{1\} + \{2\}]\psi(\alpha_1, \beta_2, \gamma') \tag{6.6}$$

which justifies the assignment of the quantum numbers α to electron 1 and β to electron 2 in calculating the matrix of any function of the type $\{1\} + \{2\}$ in LS coupling.

[23] Note that according to the convention we are using $n_1l_1n_2l_2$ indicates that electrons 1 and 2 have the *same* quantum numbers nl. A difference in quantum numbers is indicated by a prime or a superscript.

§7. The addition of an electron to an ionic multiplet

We can treat, by this same method, the problem of calculating the magnetic energy matrix for the system of states resulting from the addition of an electron to a Russell-Saunders multiplet of the next ion. The consideration of such a set of levels as thus arising from a certain "parent" multiplet is justified only in case the spread of the resulting set of levels (a triad for an added p electron, a pentad for a d) is small compared to the distances between the sets arising from different parents; otherwise we should have an "interaction between parents". The special case ^3P+s was considered by Goudsmit,[4] and the general calculation was given by Johnson[3] in his paper on two-electron systems for the case in which one adds an electron which is not equivalent to any electron in the ion. Johnson's calculation was made with the same assumptions as in the case of two-electron spectra, and requires for justification the considerations we shall now give.

Consider a Russell-Saunders multiplet of an ion. In this case the magnetic interaction in the ion is so weak that it can be considered as a small perturbation applied to the degenerate l^i, s^i energy level resulting from the electrostatic interaction. (We will use the letter i to designate resultant momenta for the ion and e to designate the quantum numbers of the added electron.) The magnetic interaction energy $V^1=\sum\xi(r_k)\boldsymbol{L}_k\cdot\boldsymbol{S}_k$ for the ion is diagonal with respect to j^i and m_j^i, and the dependence on j^i and m_j^i of each term may be obtained from Johnson's formulas (p. 1635):

$$(l^is^ij^im_j^i|\ \boldsymbol{L}_k\cdot\boldsymbol{S}_k\ |l^is^ij^im_j^i)$$
$$=(l^i|\ L_k\ |l^i)(s^i|\ S_k\ |s^i)\left\{\frac{j^i(j^i+1)-l^i(l^i+1)-s^i(s^i+1)}{2}\right\}.$$

Here the first two factors are quite independent of j^i and m_i^j. Since the matrix of $\xi(r_k)$ does not depend on j^i or m_j^i the whole dependence of the matrix of V^1 on j^i, m_j^i is given by[24]

$$(l^is^ij^im_j^i|\ V^1\ |l^is^ij^im_j^i)=\zeta\hbar^2\left\{\frac{j^i(j^i+1)-l^i(l^i+1)-s^i(s^i+1)}{2}\right\}$$
$$=\zeta(l^is^ij^im_j^i|\ \boldsymbol{L}_i\cdot\boldsymbol{S}_i\ |l^is^ij^im_j^i).$$

Hence our combination of ionic multiplet and electron is effectively a configuration l^is^i, l^es^e with a magnetic interaction energy $\zeta\boldsymbol{L}_i\cdot\boldsymbol{S}_i+\xi(r_e)\boldsymbol{L}_e\cdot\boldsymbol{S}_e$.

Now our resultant states for the atom, characterized by $LSJM_J$, are known linear combinations of states characterized by LSM_LM_S. Denote this last set of quantum numbers by γ and the eigenstate by $\Psi(\gamma)$. This latter state must be properly antisymmetrized in all electrons. Now from the functions $\phi(l^im_l^is^im_s^i;\ l^em_l^es^em_s^e)$ we can build up uniquely the state $\phi(l^is^i;\ l^es^e;\ \gamma)$ or, for short, $\psi(\alpha_1\ldots_n,\beta_e,\ \gamma)$ where in place of i we write the names

[24] This is, of course, just the Landé interval rule. Note that the numerical value of ζ can be obtained directly from the multiplet splitting of the ion.

of the electrons $1 \cdots n$ in the ion. This state is antisymmetric in all the ion electrons, but unless the quantum numbers nls of one of the ion electrons agree with those of the added electron, it can be neither symmetric nor antisymmetric with respect to interchange of the added electron with any of the ion electrons. In this non-equivalent case, we build up an antisymmetric $\Psi(\gamma)$ by taking the following linear combination

$$\Psi(\gamma) = \frac{1}{(n+1)^{1/2}}\left[\psi(\alpha_{1\cdots n}, \beta_e, \gamma) - \sum_{k=1}^{n}\psi(\alpha_{1\cdots e\cdots n}, \beta_k, \gamma)\right], \qquad (7.2)$$

where e replaces k in the subscript of α under the summation sign. Now a general term in our interaction matrix is

$$\bar{\Psi}(\gamma)\{\zeta \boldsymbol{L}_i \cdot \boldsymbol{S}_i + \xi(r_a)\boldsymbol{L}_a \cdot \boldsymbol{S}_a\}\Psi(\gamma'), \qquad (7.3)$$

where $\zeta \boldsymbol{L}_i \cdot \boldsymbol{S}_i$ acts on the ion electrons and $\xi(r_a)\boldsymbol{L}_a \cdot \boldsymbol{S}_a$ acts on the added electron.[25] Now in our case of no ion electrons equivalent to the added electron, this matrix component will vanish between states having different ion electrons, and hence a different added electron. The $(n+1)$ terms which remain will all be equal, hence (7.3) becomes simply

$$\psi(\alpha_{1\cdots n}, \beta_e, \gamma)\{\zeta \boldsymbol{L}_i \cdot \boldsymbol{S}_i + \xi(r_e)\boldsymbol{L}_e \cdot \boldsymbol{S}_e\}\psi(\alpha_{1\cdots n}, \beta_e, \gamma') \qquad (7.4)$$

which is the expression used by Johnson.

The special cases p^5x or d^9x of configurations of the type $\mathcal{L}$ may be considered as the addition of electron x to the ^2P or ^2D multiplet of the p^5 or d^9 ion. In this case since this doublet is the only multiplet of the ionic configuration the "parentage" of our states is unquestioned to the accuracy of the first order perturbation theory of Part I. The remark of footnote 24 enables us to obtain an estimate of the value of $\zeta_{n^a l^a}$ for these configurations from the doublet splitting of the ion, since the ζ of equation (7.1) equals $-\zeta_{n^a.^a}$ in this case to the approximation in which we use the same central field for the two cases.

§8. The matrix of magnetic interaction between configurations for two electrons

We will now extend the results of §6 to show how one calculates the matrix of magnetic interaction between two two-electron configurations by using the very convenient formulas of Johnson which give directly the matrix component of magnetic interaction between any two states with precise values of $l_1 l_2' s_1 s_2' LSJM_J$. This matrix is diagonal with respect to JM_J, so let us consider the matrix component joining state $SLJM_J$ of configuration $x(=\alpha\beta)$ and state $S'L'JM_J$ of configuration $y(=\gamma\delta)$.

The state $SLJM_J$ will be a linear combination of states SLM_LM_S, or say λ, while the state $S'L'JM_J$ will be a linear combination of states $S'L'M_L'M_S'$, or say λ'. The antisymmetric $\Psi(\lambda)$ will be of the form

[25] i. e., $\zeta \boldsymbol{L}_i \cdot \boldsymbol{S}_i$ acts on the electrons belonging to α, $\xi(r_a)\boldsymbol{L}_a \cdot \boldsymbol{S}_a$ on the electron belonging to β in ψ.

$$\Psi(\lambda) = \frac{1}{(2)^{1/2}}[\psi(\alpha_1, \beta_2, \lambda) - \psi(\alpha_2, \beta_1, \lambda)],$$

when we take $\alpha \neq \beta$. Similarly

$$\Psi(\lambda') = \frac{1}{(2)^{1/2}}[\psi(\gamma_1, \delta_2, \lambda') - \psi(\gamma_2, \delta_1, \lambda')],$$

when we take $\gamma \neq \delta$. The general term in the interaction matrix will be

$$\tfrac{1}{2}[\bar{\psi}(\alpha_1, \beta_2, \lambda) - \bar{\psi}(\alpha_2, \beta_1, \lambda)][\{1\} + \{2\}][\psi(\gamma_1, \delta_2, \lambda') - \psi(\gamma_2, \delta_1, \lambda')].$$

From this we easily see that we get no value for this interaction unless one of the electrons retains the same quantum numbers, say $\alpha \equiv \gamma$. In this case the remaining terms reduce to

$$\bar{\psi}(\alpha_1, \beta_2, \lambda)\{2\}\psi(\alpha_1, \delta_2, \lambda').$$

Hence we have the result that

$$\bar{\Psi}(\alpha\beta SLJM_J)[\xi(r_1)\boldsymbol{L}_1\cdot\boldsymbol{S}_1 + \xi(r_2)\boldsymbol{L}_2\cdot\boldsymbol{S}_2]\Psi(\alpha\delta S'L'JM_J)$$
$$= \bar{\psi}(\alpha_1\beta_2 SLJM_J)[\xi(r_2)\boldsymbol{L}_2\cdot\boldsymbol{S}_2]\psi(\alpha_1\delta_2 S'L'JM_J),$$

which becomes, if $\alpha = nl$; $\beta = n'l'$; $\delta = n''l''$,

$$= (n_2' l_2' \mid \xi(r_2) \mid n_2'' l_2')(l_1 l_2' SLJM_J \mid \boldsymbol{L}_2\cdot\boldsymbol{S}_2 \mid l_1 l_2'' S'L'JM_J) \qquad (8.1)$$

The value of the last factor here is given by Johnson, and in particular it is seen from his formulas to be diagonal with respect to l_2. This means that we may have a magnetic interaction only between configurations differing just in the n of one electron, the matrix being given then by (8.1).

This was for the case in which neither of the configurations was composed of equivalent electrons. We have now to consider the case in which the two electrons of configuration x are equivalent. (If both configurations are composed of equivalent electrons we get no interaction unless the states are identical, which case has already been treated.) If λ is an allowed level, $\psi(\alpha_1, \alpha_2, \lambda)$ will already be antisymmetric. The general term in the interaction matrix will then be

$$\frac{1}{(2)^{1/2}}\bar{\psi}(\alpha_1, \alpha_2, \lambda)[\{1\} + \{2\}][\psi(\alpha_1, \delta_2, \lambda') - \psi(\alpha_2, \delta_1, \lambda')]$$
$$= \frac{1}{(2)^{1/2}}\bar{\psi}(\alpha_1, \alpha_2, \lambda)\{2\}\psi(\alpha_1, \delta_2, \lambda') - \frac{1}{(2)^{1/2}}\bar{\psi}(\alpha_1, \alpha_2, \lambda)\{1\}\psi(\alpha_1, \delta_2, \lambda')$$
$$= (2)^{1/2}\bar{\psi}(\alpha_1, \alpha_2, \lambda)\{2\}\psi(\alpha_1, \delta_2, \lambda'). \qquad (8.2)$$

The balance of the computation is as above. Hence if one configuration is composed of equivalent electrons we obtain a factor of $(2)^{1/2}$ which is not obtained in the case of non-equivalent electrons. In this particular case then, one does not obtain the correct result by simply assigning definite quantum numbers to the individual electrons.

The result, that we obtain a magnetic interaction only between configurations which differ just in the n of one electron, holds quite generally for any

configurations of any number of electrons. This can be readily seen from (5.5) and (5.6). Such configurations can be considered as two members of the same series, and in general lie far apart in the spectrum. The type of series perturbation found by Shenstone and Russell[26] cannot be due to magnetic interaction, but must be purely electrostatic.

The writer intends, in the near future, to consider more in detail the question of the interaction of configurations, especially in connection with the interesting work of Shenstone and Russell on perturbed series and in an attempt to explain some of the poor agreements found by Condon and himself, references 2 and 5.

I wish to take this opportunity to express to Professor E. U. Condon my sincere appreciation of his friendly and stimulating guidance in the study of quantum mechanics in general and atomic spectra in particular.

[26] A. G. Shenstone and H. N. Russell, Phys. Rev. **39**, 415 (1932).

TERM VALUES IN THE ARC SPECTRUM OF SELENIUM*

By R. C. Gibbs and J. E. Ruedy
Cornell University

(Received February 27, 1932)

Abstract

A number of new lines in the spectrum of Se I have been observed and classified, and term values are given with respect to the lowest level of Se II, $4p^3\ {}^4S$. The first ionizing potential of the selenium atom in its lowest energy state ($4p^4\ {}^3P_2$) is 78,659 ± 2 cm^{-1} or 9.70 volts.

THE lines of the arc spectrum of selenium, classified by Runge and Paschen,[1] and appearing in Fowler's Report on Line Spectra, yield two series which determine the first ionization potential of the atom above the excited $4p^3 5p\ {}^5P_1$ level, with an error of less than one frequency number. (The terms are listed in Fowler as triplets, but should be quintets.) The term

Table I.

Intens.	λ(vac.)	λ(vac.)	Classification
12	2164.82	46193.2	$4p\ {}^3P_1 - 5s\ {}^5S_2$
15	2075.42	48183.0	${}^3P_2 - \ {}^5S_2$
15	2063.42	48463.2	${}^3P_0 - \ {}^3S_1$
20	2040.48	49008.1	${}^3P_1 - \ {}^3S_1$
20	1960.87	50997.8	${}^3P_2 - \ {}^3S_1$
8	1919.20	52105.0	${}^1D_2 - 4d\ {}^3D_1$
9	1913.81	52251.8	${}^1D_2 - \ {}^3D_2$
8	1898.57	52671.2	${}^1D_2 - \ {}^3D_3$
6	1690.72	59146.4	${}^3P_0 - \ {}^3D_1$
6	1675.28	59691.5	${}^3P_1 - \ {}^3D_1$
4	1671.18	59838.0	${}^3P_1 - \ {}^3D_2$
3	1621.26	61680.4	${}^3P_2 - \ {}^3D_1$
6	1617.40	61827.6	${}^3P_2 - \ {}^3D_2$
6	1606.51	62246.7	${}^3P_2 - \ {}^3D_3$
2	8054.2	12415.9	$5p\ {}^5P_1 - 7s\ {}^5S_2$
4	8083.6	12370.7	${}^5P_2 - \ {}^5S_2$
4	8153.0	12265.4	${}^5P_3 - \ {}^5S_2$
3	8921.5	11208.8	$5s\ {}^5S_2 - 5p\ {}^5P_3$
2	9004.5	11105.4	${}^5S_2 - \ {}^5P_2$
1	9041.3	11060.3	${}^5S_2 - \ {}^5P_1$

value of the $4p^3 5s\ {}^5S_2$ level given there was computed from a Rydberg formula, the second member only of the ${}^5S - {}^5P$ series having been observed. The first member of this series has now been observed by us (near λ9000) and the term value of the ${}^5S_2$ level thus definitely fixed. This now permits the levels estab-

* This research was supported by a grant from the Heckscher Foundation for the Advancement of Research, established at Cornell University by August Heckscher.

[1] C. Runge and F. Paschen, Astro-Phys. Jour. 8, 70 (1898).

lished by McLennan, McLay and McLeod,[2] from observations in the vacuum region, to be given term values with respect to the limit determined by the Runge-Paschen series.

During our investigation the entire region from λ1,000 to λ10,000 was photographed on various grating and prism instruments and many new lines

TABLE II. *Se I Terms.*

	$4p$	$^3P_2$:78659 $^3P_1$:76670 $^3P_0$:76125	$^1D_2$:69083	$^1S_0$:55289 (McLennan[3])
		$4s^23d^{10}4p^3:{}^4S$	$4s^23d^{10}4p^3:{}^2D$	$4s^23d^{10}4p^3:{}^2P$
	$5s$	$^5S_2$:30476 $^3S_1$:27661	^1D ^3D	^1P ^3P
	$5p$	$^5P_1$:19416 $^5P_2$:19371 $^5P_3$:19268 $^3P_{012}$	$^1(PDF)$ $^3(PDF)$	$^1(SPD)$ $^3(SPD)$
$4s^23d^{10}4p^3$	$4d$	$^5D_{01234}$ $^3D_1$:16979 $^3D_2$:16832 $^3D_3$:16412	$^1(SPDFG)$ $^3(SPDFG)$	$^1(PDF)$ $^3(PDF)$
	$6s$	$^5S_2$ $^3S_1$		
	$6p$	$^5P_1$:9393 $^5P_2$:9379 $^5P_3$:9342 $^3P_{012}$		
	$5d$	$^5D_{01234}$: $^3D_{123}$		
	$7s$	$^5S_2$:7001 $^3S_1$		
	$6d$	$^5D_{01234}$:5112 $^3D_{012}$		
	$8s$	$^5S_2$:4449 $^3S_1$		

observed, some of which fit into the term scheme so far developed, but many of which do not. Most of the new lines which fit, lie in the vacuum region, and our values for those observed first by McLennan are somewhat more accurate than his. A six-foot grating vacuum spectrograph was used in this region and the spectrum was excited by a spark between cored aluminum electrodes. At the longer wave-lengths, two six-foot gratings in Rowland mountings were used for accurate measurements, and an electrodeless discharge, as well as a vacuum spark, was used for the excitation of the selenium.

In a letter to Nature, McLennan and Crawford[3] refer to the $4p^4\ ^1D_2$ and

[2] J. C. McLennan, A. B. McLay, and J. H. McLeod, Phil. Mag. **4**, 486 (1927).
[3] J. C. McLennan and M. F. Crawford, Nature **124**, 874 (1929).

$4p^4\ {}^1S_0$ levels of selenium and give their positions with reference to the low $4p^4\ {}^3P_{210}$ levels. Our data confirm the ${}^1D_2$ assignment but neither confirm nor contradict that of the ${}^1S_0$.

A list of all the classified ultraviolet lines and other new lines which have been classified is given in Table I.

Table II gives all the known terms, starting from the deepest up to and including the first of the ${}^5S$ and ${}^5D$ terms tabulated in Fowler.

The investigation is being continued.

THE CALCULATION OF ERRORS BY THE METHOD OF LEAST SQUARES

By Raymond T. Birge
University of California, Berkeley

(Received February 18, 1932)

Abstract

Present status of least squares' calculations.—There are three possible stages in any least squares' calculation, involving respectively the evaluation of (1) the most probable values of certain quantities from a set of experimental data, (2) the reliability or probable error of each quantity so calculated, (3) the reliability or probable error of the probable errors so calculated. Stages (2) and (3) are not adequately treated in most texts, and are frequently omitted or misused, in actual work. The present article is concerned mainly with these two stages.

Validity of the Gaussian error curve.—All least squares' calculations of probable error assume that the residuals follow a Gaussian error curve. This curve is derived from a consideration only of *accidental* errors. Probable errors are, however, evaluated frequently in cases where constant or systematic errors are known to be present. Such a procedure, when used judiciously, is believed by the writer to be better than any alternative procedure, but the results are naturally less reliable than a strict reliance on theory would indicate. The statement is sometimes made that in practise one often gets more *large* residuals than are predicted by theory. This point is tested by 500 measurements of a spectral line, and the resulting Gaussian curve, plotted in Fig. 1, shows no indication of such a deviation. It is possible to account for an excess of large residuals (with the necessary accompaniment of a deficiency of small residuals), by assuming that the various observations used were not in fact of equal reliability. The formulas by which the probable error r may be calculated from the observed residuals, and the reliability of each such value, are briefly considered.

Internal versus external consistency.—Probable errors are calculated usually on the basis of internal consistency, and most texts discuss only this method. Scarborough, in a recent article, claims that there is no logical basis for the calculation from external consistency. This matter is considered in detail, and it is shown that the two methods must necessarily lead to the same result, except for statistical fluctuations, provided that only accidental errors are present. Formulas for the magnitude of the expected fluctuations are given. It is shown that a probable error based on internal consistency (r_i) is virtually a *prediction* and that the probable error based on external consistency (r_e) is the answer to this prediction. When the ratio r_e/r_i exceeds unity by an amount much greater than is to be expected on the basis of statistical fluctuation, one has almost certain evidence of the presence of systematic errors. In such a case new arbitrary weights should be assigned. Then one has available *only* external consistency as a basis for the calculation of errors. The false deductions that result from a failure to note the above facts are illustrated by several examples from the literature, and a numerical problem illustrating all of these relations is presented in detail.

Probable error of a function evaluated by least squares.—The probable error of a function of directly observed quantities is given by the well-known law of "propagation of Error." When, however, one has a function whose coefficients have been evaluated by least squares, the probable error of the function is scarcely mentioned in the literature, and the writer has never seen it calculated in practise. Explicit formulas for the error of a rational integral function of any degree are derived by the writer, from elementary considerations, and are found to agree with the very general formula already known. Practically all of the results presented here find frequent application in the article on probable values of e, h, etc., which immediately follows.

Introduction

THE method of least squares furnishes a precise and reliable process of calculation in much scientific work. That its value is not more generally appreciated seems to be due to the fact that it is often misused or only incompletely used, with resulting discredit to the system as a whole. There are three possible stages in any least squares' calculation, (1) the calculation of the "most probable" values of certain quantities, from a given set of experimental data, (2) the calculation of the "probable error" of each of the quantities just evaluated, (3) the calculation of the reliability, or probable error, of the probable errors so evaluated.

The proper procedure for the first stage is covered quite adequately in almost any good text on the subject. The second stage is covered much less adequately, and certain formulas that it appears to the writer should find frequent use in scientific work have apparently never been stated explicitly and seem never to have been used. The general theory covering these two stages has, however, been known for a long time, so that the working out of explicit formulas is a matter of mere detail. In the case of the third stage, the theory is very incomplete and rather uncertain, even in its simpler aspects.

The present article is concerned chiefly with the second and third stages of the calculation, since most of the misapplication of least squares in the literature has been in connection with these matters. Least squares is a branch of statistical theory and it would seem that there is in statistics a certain elusive character that makes all deductions difficult to state clearly. Certainly the writer must confess that he has consulted practically every available text on least squares, and yet has had considerable difficulty in understanding many points. The numerous instances in the literature of the injudicious or incorrect use of least squares seem to indicate that others have had similar difficulties. This must be the excuse for the discussions to follow, which, at first glance, may appear to be presented in quite unnecessary detail. Most of the facts may be found in any good text, but some of the formulas and points of view are, I believe, new. My object has been to bring out what may crudely be termed the "physics" of the matter, rather than the mathematics. In fact practically no mathematical proofs are given, since these may be found in any number of texts.

The direct incentive to the writing of this article has been two-fold, (1) the publication of a paper by Scarborough[1] entitled "The Invalidity of a Commonly Used Method for Computing a Certain Probable Error," and, (2) the need of certain formulas in my paper[2] on "Probable Values of e, h, e/m and α," which follows the present article. Scarborough's article considers a matter that is ignored in practically all texts, but his main conclusion is unfortunately quite incorrect. In my own article I have had occasion to use formulas that, so far as I know, are not stated in convenient working form

[1] J. B. Scarborough, Proc. Nat. Acad. Sci. **15**, 665 (1929). This article is given, practically verbatim, in his book "Numerical Mathematical Analysis" (Johns Hopkins Press, 1930) pp. 328–332.

[2] This paper on e, h, etc. will hereafter be denoted as G.C. 1932.

anywhere in the literature. The present paper contains the derivation of these formulas, and G.C. 1932, which follows, may be considered as an illustration of nearly every point discussed here.

The Gaussian Error Curve

The first important fact concerning the method of least squares is that it applies strictly only to sets of measurements whose errors follow the Gaussian error curve. It may be proved that this law should be followed if the errors are purely accidental (not systematic or constant), are equally liable to be positive or negative, and are more likely to be small than large, with very large errors entirely lacking.[3] Taken too literally, this requirement would practically eliminate least squares as a method of computation, for in nearly every set of measurements there is presumably a certain amount of systematic or constant error. The probable error calculated by least squares takes account only of accidental errors, but it is always possible to enlarge this probable error to make what seems proper provision for other types of errors. Let us suppose that a given constant has been measured in several different ways, and that the stated probable error of each result makes allowance for the various possible sources of error. If now there is no reason to believe that the errors in one investigation are related, in sign or magnitude, with those of another investigation, it seems quite legitimate to combine by means of least squares the various results, i.e., to weight them according to their stated probable errors, and to derive the probable error in the final weighted average by the usual formulas. This is a very important point which will appear later in the paper and to which some readers may take exception. We assume, in essence, that the constant or systematic errors occurring in a series of unrelated investigations also follow a Gaussian error curve. It must be admitted that in practise one is very unlikely to meet such an ideal distribution of constant errors, and for that reason the final result and its probable error must be considered *less* reliable than the theory of least squares would indicate. What I do maintain is that such a procedure has more to recommend it than any alternative procedure and that, if used with proper discretion, it will result in no serious error. One may summarize the situation by saying that, except in especially favorable cases, least squares' results, and their computed probable errors, are *not* as reliable as indicated by theory. This fact seems unfortunately so well known that many persons have chosen to use other methods (or no method at all!) for calculating their results. I use the word "unfortunately" because these alternative methods are, without exception, inferior to least squares. Certainly one *must* make computations, in scientific work, and the fact that a certain system of computation is not as reliable as over-zealous advocates may claim is no excuse for using in its place a still less reliable system.

[3] See, for instance, M. Merriman "Method of Least Squares" 8th edition, articles 17–28, or Whittaker and Robinson "Calculus of Observations" pp. 168–173. The validity of the proposed proofs of the Gaussian error curve has been the subject of endless discussion, but this article is not concerned with the philosophical aspects of least squares.

The equation of the Gaussian error curve is

$$y = \frac{h}{(\pi)^{1/2}} e^{-h^2 x^2} \tag{1}$$

where y is the proportional number of observations having an error x, and h is the so-called precision index. The curve applies strictly only to errors, but these are never known, since the true magnitude of the measured quantity is also not known. All that we know is the most probable magnitude of the quantity, or what is often termed its adjusted value, and the deviation of any measurement from this is called the residual v. If a sufficient number of measurements are available to make at all possible the plotting of an error curve, the residuals should follow the same equation and lead to essentially

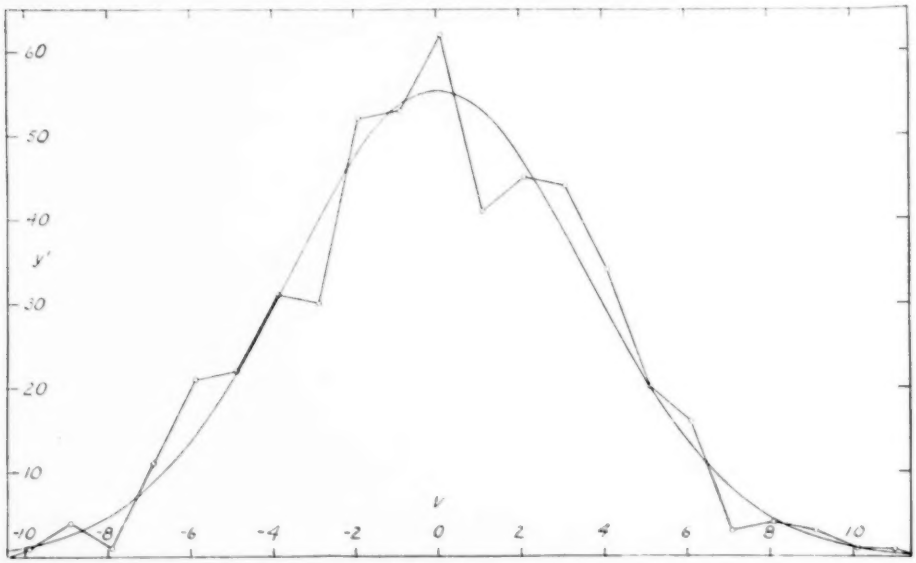

Fig. 1. The distribution of residuals for 500 measurements of a spectral line, compared with the Gaussian error curve evaluated by second moments. Abscissa represents residual v, in 0.001 mm units. Ordinate represents the number of residuals of magnitude v.

the same calculated value of h. In other words, the difference in the results given by errors and by residuals should be negligibly small, and in the following discussion this is assumed to be the case. Now in most scientific work only a comparatively small number of observations are taken, and it is manifestly impossible to make a direct test of the distribution of the residuals. Yet least squares' formulas are applied to these small sets of observations. In such cases the following implicit assumption is made. One assumes, from his own previous experience or from that of others, that if the number of observations were greatly extended, with no change in conditions, it would be found that this large set agreed with the Gaussian error curve.

The statement is sometimes made[4] that in such extensive sets of measure-

[4] See, for instance, Tuttle and Satterly "Theory of Measurements" p. 175. The Frontispiece of G. C. Comstock "Method of Least Squares" shows four error curves, and three of these

ments, one generally finds more large residuals than are predicted by theory. In order to test this statement, I have made a series of 500 cross-hair settings on a very wide[5] but symmetrical solar spectrum line, under conditions as favorable as possible to equal reliability for all observations. Readings were made to the nearest 0.001 mm, and the arithmetic average was adopted as the most probable value. The residuals calculated with respect to this average are scattered over 12 positive and 10 negative values. The resulting distribution is shown in Fig. 1, in which the unit abscissa is 0.001 mm and the plotted ordinate is y' $(=500y)$. y' thus represents the actual number of residuals of a given magnitude.

In fitting these points to Eq. (1), the one adjustable constant h is calculated by moments, normally by second moments.[6] From the definition of y we have as alternative expressions for the second moment of the points, $\Sigma yx^2 = \overline{x^2} \cdot \Sigma y = \overline{x^2} = \Sigma x^2/n$, where the last summation is to be extended over the entire n original observations. The second moment of the calculated curve, $\int_{-\infty}^{+\infty} yx^2 dx$ equals [7] $1/2h^2$. Hence, by equating the two second moments we obtain

$$\frac{1}{h} = \left(\frac{2}{n}\sum x^2\right)^{1/2}. \tag{2}$$

If the errors x are replaced by the residuals v, Eq. (2) becomes[8]

$$\frac{1}{h} = \left(\frac{2}{n-q}\sum v^2\right)^{1/2} \tag{3}$$

where q is the number of constants simultaneously evaluated by the least squares' solution. In the case of a series of observations of a single quantity, $q=1$.

---

do show a slight indication of this excess of large residuals. On the other hand, Merriman, reference 3, page 33, presents 300 astronomical observations that agree very well with the calculated error curve, and he remarks "Whatever may be thought of the theoretical deductions of the law of probability of error, there can be no doubt but that its practical demonstration by experience is entirely satisfactory". Scarborough "Numerical Mathematical Analysis", page 314, gives 470 astronomical observations that show an equally good agreement with theory.

[5] The width was achieved in part by the use of high magnification in the comparator telescope.

[6] The term second moment comes from the close analogy to mechanics. (See T. C. Fry, "Probability and Its Engineering Uses", p. 183). If all the observed points in the negative quadrant of Fig. 1 are folded over onto the positive quadrant, and each point is given a mass equal to its y value, then the second moment of the observations, Σyx^2, is merely their moment of inertia with respect to the y axis. Similarly, if the calculated curve encloses a flat body of constant mass per unit area, such that the total mass of this body equals the total mass of the points, then $\int yx^2 dx$ represents its moment of inertia. In the calculation of h by second moments one equates $\int yx^2 dx$ to Σyx^2 and thus effectively makes the two moments of inertia equal. Since the masses are necessarily equal, one may say that the "radius of gyration" of the calculated curve is equated to the radius of gyration of the points. Similarly, in the case of *first* moments, the "lever arm" of the force moment of the points is equated to the corresponding lever arm of the calculated curve.

[7] See Merriman, reference 3, page 73.

[8] See Merriman, reference 3, pp. 79–82.

The probable error r of one observation (i.e., the value of $|x|$ that divides the Gaussian curve into two equal areas) is related to h by $r=0.4769/h$, so that

$$r = 0.6745\left(\frac{\Sigma v^2}{n-q}\right)^{1/2}. \tag{4}$$

If h is calculated by first moments (by equating $\int yxdx$ to Σyx), one obtains

$$r = 0.8453\frac{\Sigma|x|}{n} \tag{5}$$

which is known as Peter's formula.

Gauss was able to show[9] that the value of h, or of r, when calculated by second moments is more reliable than that calculated by moments of any other degree. His general formula gives for the proportional probable error in h, or in r, when obtained by second moments,

$$u = \frac{0.4769}{(n)^{1/2}}. \tag{6}$$

If r is calculated by first moments, the numerator in Eq. (6) becomes 0.5096, if by third moments 0.497, etc.

This seems to be the extent of the theory, so far as concerns the third stage of the calculations, mentioned in the introduction. The derivation of Eq. (6) was made by Gauss in connection with the curve of *errors*, and it is still uncertain[10] whether n or $n-1$ should be used in actual calculations with residuals. It seems to me that when one is fitting a curve containing q constants, so that the probable error of an observation of unit weight[11] is given by Eq. (4), one should use $n-q$ in place of n, in Eq. (6). I am not, however, certain about this, and I shall therefore continue to use Eq. (6), as I have done in the past. It seems safe to conclude, in any case, that the value of u, aside from other sources of uncertainty, is unreliable to the amount calculated by replacing n by $n-q$ in Eq. (6).

The smooth curve drawn in Fig. 1 corresponds to $h=0.1965$, the value calculated from second moments. By Eq. (6) its uncertainty is 2.13 percent. The value of h from first moments is 0.1959, a change of only 0.3 per cent, and the reliability of this latter value is 2.28 percent. These two methods for computing h thus yield almost identical results, in this case, and always have almost equal reliability. The two calculated curves agree too closely to be conveniently shown together in Fig. 1.

This figure certainly gives no evidence of an excess of large residuals, and in fact the entire course of the observations follows the Gaussian curve with

[9] See Whittaker and Robinson, reference 3, p. 201.

[10] See Merriman, reference 3, p. 208.

[11] The question of weighting, and the meaning of "unit weight" is considered later in this paper.

surprising accuracy.[12] The one noticeably "low" point at $v = +1.1$, bordered by the "high" point at $v = +0.1$ is in all probability due to a slight irregularity of the comparator screw, which causes one setting to be favored at the expense of an adjacent setting. This incidentally furnishes a good illustration of the fact that errors, other than accidental, are *always* present to greater or less degree, in actual scientific work. I believe that the apparent excess of large residuals sometimes noted in practise is due usually to the unsuspected presence of observations of unequal reliability. If a portion of the observations has actually one precision index h, and another portion has another, it is easy to show that a single plot of all the residuals *should* exhibit an apparent excess of large residuals. Now the area under the observed curve[13] and under the calculated Gaussian error curve is necessarily the same (unity for each). Hence it follows that any observed excess of large residuals must be compensated by a numerically equal deficiency of small residuals. In the case of the large residuals, however, the *proportional* excess is large, while the proportional deficiency of small residuals is small. Thus the observed excess stands out, while the correlated deficiency may easily pass unnoticed.[14]

Internal Versus External Consistency

Scarborough's article[1] is concerned with the relative merits of internal and external consistency, in the calculation of probable errors, although he does not employ these particular designations. This is a matter that is very inadequately treated in text-books, and concerning which there appears to be wide-spread misunderstanding. It is accordingly presented in considerable detail.

Let us assume that we have made n direct observations, each of the

[12] N. R. Campbell, on page 182 of his book "Measurement and Calculation", remarks concerning published distributions of residuals, "The examples offered are never typical of those occurring in physics (of other sciences, such as astronomy or geodesy, I say nothing). In my experience, the residuals in physical examples usually depart notably from the law, the curve representing them having a flatter top than the Gaussian. This is what our theory would predict; for there is no reason to believe that all errors considerably less than the maximum have not equal chances."

The last sentence is quite correct. By *sufficiently* restricting the range of x, the Gaussian error curve is itself essentially "flat". The data plotted in Fig. 1 are typical of the sort of work on which the writer is engaged. Whether or not it is physics may be left to the reader. In this connection it should be noted that Campbell advocates, in place of least squares, the "method of averages," discovered by Tobias Mayer in 1748 and in common use by statisticians at the present time (see Whittaker and Robinson, reference 3, pp. 258–259, H. L. Rietz "Handbook of Mathematical Statistics", p. 64, and Scarborough, reference 1, pp. 357–363). This method was "rediscovered" by Campbell in 1920 (N. R. Campbell, Phil. Mag. **39**, 177, 1920), and named by him the "method of zero sum", a name he still employs (Phil. Mag. **10**, 745, 1930). In 1924 he remarked (Phil. Mag. **47**, 816, 1924), "so far as I know, the method is not habitually used by any one but myself".

[13] If, as in Fig. 1, the observed points are connected by straight lines, ending on the v-axis as shown, the area under this series of lines (the "observed" curve) is n, if y' is plotted as in Fig. 1, or unity if the proportional number of residuals y is plotted. The total area under the calculated Gaussian curve, Eq. (1), is of course unity.

[14] This is very well illustrated by the curves in Comstock's text, mentioned in footnote 4.

same reliability, of a certain quantity, and have calculated the arithmetic average and the deviations (residuals) from this average. Then, by Eq. (4), the probable error in any one observation is

$$r = F\left(\frac{\sum v^2}{n-1}\right)^{1/2} \tag{7}$$

where[15] $F = 0.6745$. This equation is essentially a statement of certain facts. As has been brought out in the preceding discussion, it is assumed that the residuals actually follow a Gaussian error curve, or would do so if the observations were sufficiently numerous. From the observed distribution of the residuals we then deduce, by means of Eq. (7), that particular residual r that divides the area under the Gaussian curve into two equal parts, since this equation expresses the most reliable relation between r and the residuals v.

From this observed value of the probable error of a single observation, we now proceed to predict the probable error in the arithmetic average. This is done by applying the general equation of the Propagation of Errors (see Eqs. (24) and (25) ahead). The result is

$$(R)_{int} = F\left(\frac{\sum v^2}{(n-1)n}\right)^{1/2}. \tag{8}$$

I think that the most important requirement in the present discussion is a recognition of the essential difference between Eqs. (7) and (8). In the case of Eq. (7) we have a number of residuals, and as just indicated, this equation is merely a compact and useful expression for the distribution of these residuals. We have, however, only *one* arithmetic average and Eq. (8) is therefore a direct prediction. It states in effect the following. Let us imagine that instead of *one* set of n observations of a certain constant reliability, we have N such sets, each composed of n observations of this same reliability. We may then calculate the N arithmetic averages, which will hereafter be called *points*, and these points, just as in the case of the n observations of any set, should disagree more or less with one another. If the arithmetic average of the set of points is calculated, and also the deviations V of each point from the average, this new set of residuals should in turn follow a Gaussian error curve, for which the probable error is[16]

[15] This nomenclature for the factor 0.6745 is introduced merely to save type-setting.

[16] A great deal of my own earlier confusion regarding internal and external consistency has been due to ambiguous nomenclature. To get the facts clearly in mind, it seems necessary to make a sharp distinction between the *points* on a curve and the *observations* of which each point is the average. To aid in drawing this distinction, I have used lower case type for all letters connected with the observations (such as v, n, r) and capitals for letters concerned with points. In the next section of this paper, and in general work, no such distinction seems necessary, and accordingly in G.C. 1932 and elsewhere I use the customary lower case letters for both points and observations.

$$(R)_{ext} = F\left(\frac{\sum V^2}{N-1}\right)^{1/2}. \tag{9}$$

Eq. (9) is thus the *answer* (actual or hypothetical) to the *prediction* made by Eq. (8). $(R)_{int}$ is the probable error of a point, as calculated by *internal* consistency, and $(R)_{ext}$ is the probable error of this same point, or any other of the N points, as calculated by *external* consistency.

If now the experimental conditions are such as to justify the application of the theory of least squares, on which the predicted value $(R)_{int}$ is based, the two values $(R)_{ext}$ and $(R)_{int}$ should agree, *except for statistical fluctuations*.[17] In the simplified situation assumed here, it is easy to show just what the equality of $(R)_{ext}$ and $(R)_{int}$ indicates.[18] In Eq. (8) $\Sigma v^2/(n-1)$ actually represents a certain *average* value of e^2, where e is the true error of an observation,[19] i.e.,

$$\frac{\sum v^2}{n-1} = \overline{e^2}. \tag{10}$$

Similarly, in Eq. (9),

$$\frac{\sum V^2}{N-1} = \overline{E^2}. \tag{11}$$

Let us now form the ratio of $(R)_{ext}$ to $(R)_{int}$. From Eqs. (8), (9), (10) and (11) we get

$$\frac{(R)_{ext}}{(R)_{int}} = \left(\frac{\overline{E^2}}{\overline{e^2}/n}\right)^{1/2}. \tag{12}$$

Hence the statement that $(R)_{ext}$ should equal $(R)_{int}$, except for statistical fluctuations, is merely the statement that the root-mean-square average of the errors of the N points should be only $1/(n)^{1/2}$ as large as the root-mean-square average of the errors of the n observations of any set. But this is exactly the relation predicted by Propagation of Errors, and already used in deriving Eq. (8) from Eq. (7).

Before discussing the expected variation from unity of the $(R)_{ext}/(R)_{int}$ ratio, let us consider a more general situation. We assume now that the N sets of observations from which the N points are calculated, are formed of varying numbers n_i of individual observations, and that the reliability of the observations also varies from one set to another. In that case we calcu-

[17] The expected magnitude of such fluctuations will be considered presently.

[18] A. Palmer "Theory of Measurements" pp. 66–71, gives a more abstract discussion of the equality of these two probable errors. This is the only reference in English to this important question that seems worthy of quotation.

[19] This has been pointed out in connection with Eqs. (2) and (3). Theory gives directly the value of r (or of h) in terms of the errors e. The theory further indicates that if the observable Σv^2 is substituted for the hypothetical Σe^2, the *most probable* relation is $\Sigma e^2/n = \Sigma v^2/(n-1)$ in the case of one unknown quantity, or $\Sigma v^2/(n-q)$ in the case of q unknowns.

late the probable error of *each* point by Eq. (8), and denote any representative point by $A_k \pm R_k$, where

$$R_k = F\left(\frac{\sum v^2}{(n_k - 1)n_k}\right)^{1/2} \cdot \cdot \tag{13}$$

This, it should be noted, is the probable error of the point, by *internal* consistency, and will henceforth be called $(R_k)_{int}$.

We next desire to combine together the N points, and it is now not necessary to assume that these N points are measurements of one single quantitv We shall, instead, assume that the N points are connected by some functional relation, i.e., that they are the values of y, for a certain set of N values of x, in $y=f(x)$. In the previous example, the N points lay on the simplest of all curves, the horizontal line $y = constant$.

Before proceeding to combine our N points, we must first *weight* each point according to its probable error, $(R_k)_{int}$, assuming that this probable error is confined *solely* to the measurement of the y coordinate.[20] Now the probable error of a measured quantity may be termed an absolute measure of its reliability. The weight of such a quantity, on the other hand, is only a relative measure of its reliability. Moreover, the weight P is defined in such a manner that it is not directly proportional to the probable error R but is connected with it by the following relation,[21]

$$R = \left(\frac{C}{P}\right)^{1/2} \quad \text{or} \quad P = \frac{C}{R^2} \tag{14}$$

where C is an arbitrary constant. Furthermore, in Eq. (13), the reciprocal of the term $\Sigma v^2/(n_k-1)$ may be termed a measure of the reliability of a single observation, as shown in connection with Eq. (10), while n_k measures the number of observations. Hence for observations of equal reliability, the probable error of the arithmetic average varies as $1/(n)^{1/2}$, and by Eq. (14), the probably error varies also as $1/(P)^{1/2}$. Thus the weight of the arithmetic

[20] This assumption of zero error (or infinite weight) in the x coordinate is made implicitly in the case of every equation presented in this article. Fortunately the condition is actually fulfilled in many of the measurements of physical science, such as the plotting of spectral frequencies against the values of a quantum number. Moreover, when all observations have the same probable error in the y coordinate, and a similarly constant error in the x coordinate, it is practically immaterial how these errors compare with one another. In particular, one may assume (1) zero error in the x coordinate, and any constant probable error in the y coordinate, or (2) zero error in the y coordinate and any constant error in the x coordinate. I have found, with typical sets of data conforming to a linear equation, that the values of the constants given by solutions (1) and (2) differ by only about one-fifth of the probable errors of the constants. Such differences are of course devoid of any real significance. On the other hand, if the probable errors for either the x or the y coordinates are not constant, solutions (1) and (2) may be appreciably different. In the case of a non-linear equation, such a difference is likely to exist whether the probable errors are constant or not. I am indebted to Dr. W. E. Deming for the foregoing general conclusions, which he has obtained in the course of a private correspondence on this subject. The basic equations on which they rest appear in his paper, Phil. Mag. **11**, 146 (1931).

[21] Palmer, reference 18, article 44. Merriman, reference 3, articles 41–43, 61 and 63.

average of n equally reliable observations is proportional to n, and if each observation is arbitrarily assumed to have unit weight, then the weight of the arithmetic average is said to be n. This is the real origin of the idea of weight.

The most explicit expression for the weight P is obtained by substituting for R in Eq. (14) its value from Eq. (13). We then get

$$P = \frac{C}{F^2}\left(\frac{n-1}{\sum v^2}\right)n. \tag{15}$$

This expression states that the weight of an arithmetic average is directly proportional to the number of observations composing it, and also to the reliability of the individual observations.

Thus, from Eqs. (13) and (14), we may calculate a weight P_k for each point from its (internal consistency) probable error $(R_k)_{int}$. Now in many actual situations we do not know the details of the derivation of the values $(R_k)_{int}$. Furthermore it is possible, as discussed earlier, that this probable error is an estimate that includes not only the effect of accidental errors, Eq. (13), but also of other possible types of error. In that case $(R_k)_{int}$ is merely a stated quantity, and is connected with other quantities of interest to us only by Eq. (14).

We shall denote by R_i the probable error of a "point of unit weight."[22] Then by Eq. (14)

$$R_i = (C)^{1/2}. \tag{16}$$

Hence

$$(R_k)_{int} = \left(\frac{C}{P_k}\right)^{1/2} = \frac{R_i}{(P_k)^{1/2}}. \tag{17}$$

Just as in the previous illustration, the value $(R_k)_{int}$ is one based on certain internal evidence and is essentially a prediction. To check this prediction we proceed to fit the N points to the known functional relation[23] $y=f(x)$. This is to be carried out by means of the appropriate least squares' formulas, that is, the q undetermined constants of the function are to be thus evaluated. One then calculates values of y and compares these with the observed values, thus obtaining a set of residuals V. Then the probable error of any point, of weight P_k, as based on *external* consistency, is given by

$$(R_k)_{ext} = F\left(\frac{\sum PV^2}{(N-q)P_k}\right)^{1/2}. \tag{18}$$

[22] This expression may easily be very confusing. It does not refer to any actual point. The constant C is purely arbitrary, and R_i is merely its square root. The magnitude of R_k for any actual point is of course *not* arbitrary. From the real magnitudes R_k and the arbitrary constant C, or R_i, one derives the weights P_k whose absolute values are arbitrary, but whose relative values are not.

[23] In order that our comparison of the values of R_k, based on internal and on external consistency, may be valid, it is necessary that the points actually be connected by the assumed functional relation. Sometimes, as in the case where the points are merely repeated measurements of a given quantity, there is no doubt as to the function. In other cases there may be considerable uncertainty on this matter.

It is convenient, in analogy with Eq. (16), to define the probable error, based on external consistency, of a point of unit weight. This, by Eq. (18) is

$$R_e = F\left(\frac{\sum PV^2}{(N-q)}\right)^{1/2}. \tag{19}$$

Eq. (18) can then be rewritten as

$$(R_k)_{ext} = \frac{R_e}{(P_k)^{1/2}}. \tag{18'}$$

It is possible to calculate also the probable errors of the q constants whose magnitudes have been simultaneously evaluated by the least squares' solution, and of the function itself. To do this we need only the *weight* of each of these quantities, and formulas for these weights will be given later. We are here, however, interested primarily in a comparison of external and internal consistency, and in making such a comparison it is immaterial whether we discuss some particular point on the curve, a hypothetical point of unit weight, one of the constants of the function, or the function itself. This is due to the fact that, as shown by Eqs. (17) and (18'), the appropriate weight of any one of these quantities enters in the same manner into both the external and the internal consistency error, and therefore cancels out when the ratio is taken. We shall, for convenience, consider the hypothetical point of unit weight. Then from Eqs. (16) and (19)

$$\frac{R_e}{R_i} = F\left(\frac{\sum PV^2}{(N-q)C}\right)^{1/2}. \tag{20}$$

This ratio, just as in the case of Eq. (12), should equal unity, except for expected statistical fluctuations. Eq. (20) appears, at first glance, to be quite different in form from Eq. (12), but it is actually only a more general expression.[24]

Since R_i by Eq. (16) is only an arbitrary constant, the uncertainty in the ratio R_e/R_i depends wholly on the uncertainty in R_e. Now the uncertainty in the calculation of the probable error of N points (or observations) is given by Eq. (6).[25] Thus if $N=9$, we may expect that the ratio R_e/R_i of Eq. (20) has itself a probable error of about 16 percent, and therefore we may expect a deviation from unity of this order of magnitude.

[24] This may be seen more clearly by substituting for each P, in the ΣPV^2, its value C/R^2 from Eq. (17). The arbitrary C then cancels out of Eq. (20) and we have under the radical the various $(R_k)_{int}$ which, by Eq. (13), presumably involve the residuals v of the original observations. The right side of Eq. (20) equated to unity, then expresses the general expected relation between the magnitude of the residuals (or errors) of the various sets of observations forming the points, and the magnitude of the residuals (or errors) of the various points, just as Eq. (12) does in one very special case.

[25] The residuals of the N points, from the calculated $f(x)$, should of course follow the Gaussian error curve, regardless of the particular function. The n observations of a single quantity, in connection with which Eq. (6) has been stated, satisfy the particular function $y = constant$.

Let us suppose now that, in a certain situation, the actually calculated ratio differs from unity by several times—say five times—the probable error of Eq. (6). There is approximately only one chance in a thousand that such a deviation is the result of mere statistical fluctuation. We therefore conclude that a definite discrepancy exists between the results based on external and on internal consistency. This discrepancy can be due to only two possible causes, (1) the functional relation to which the points have been fitted is not the true one, (2) the probable errors that have been assigned to at least a portion of the points are false.

The present discussion is not concerned with (1), and we here merely assume that the true functional relation is known. Cause (2) is due, nearly always, to the presence of unsuspected constant or systematic errors in a portion or in all of the points. The evaluation of the ratio R_e/R_i is thus an invaluable aid to the detection of such errors. If now it has been proved, from the value of this ratio, that the stated probable errors of the points are false, then the weights P_k, based on these assigned errors, are entirely misleading. In such an annoying situation, the logical procedure is to discard this set of weights, and to make an arbitrary assignment. Such an assignment is necessarily based on the personal judgment of the computer, and may be very uncertain. It is, however, likely to be better than an assignment that has been proved false. Using the new set of weights, one then makes a new calculation of the constants in $f(x)$, derives a new set of residuals V, and with these calculates a new set of probable errors by means of Eq. (18). In this situation one has available *only* external consistency for the calculation of the errors.

If the ratio R_e/R_i does not differ from unity by more than the expected amount, one may use *either* R_i or R_e for the calculation of errors, i.e., either Eq. (17) or Eq. (18′), assuming P_k to represent the appropriate weight of the quantity discussed. In the following article, G.C. 1932, I have adopted the conservative policy of using, in such a situation, the *larger* of the two quantities R_e and R_i. One often encounters the situation in which the number of points N is relatively small. The probable error in R_e is then large, and most writers consequently advocate using R_i. In fact the probable error based on internal consistency is always more reliable than one based on external consistency.[26] An additional advantage in the use of $(R_k)_{int}$ is that it can be calculated far more rapidly than $(R_k)_{ext}$. It must be remembered, however, that this conclusion as to the relative reliability of internal and external consistency depends wholly on the correctness of the assumption that only accidental errors are present, and I again emphasize the point that the

[26] The reliability of any probable error $(R_k)_{int}$ must be judged from its origin. If this probable error is given by Eq. (13), then its reliability is given by Eq. (6) with $n = n_k$. This is also the reliability of $(P_k)^{1/2}$ since in Eq. (17) R_i is merely a constant. On the other hand, the reliability of $(R_k)_{ext}$ is affected not only by $(P_k)^{1/2}$ in Eq. (18), but also by ΣPV^2. The ratio $(R_k)_{ext}/(R_k)_{int}$ is affected merely by the uncertainty in ΣPV^2, and this is given by Eq. (6) with $n = N$. Hence $(R_k)_{ext}$ is always less reliable than $(R_k)_{int}$ and it is apparently for this reason that $(R_k)_{int}$ is commonly employed in the literature.

sole reason for calculating $(R_k)_{ext}$, in addition to $(R_k)_{int}$, is to test this assumption which, unfortunately, will very often be found quite untenable.

The final conclusion of this section is that in the situation where no appreciable systematic errors are present, as shown by a satisfactory value of the ratio R_e/R_i, one may use either external or internal consistency as a basis for the calculation of probable errors. The internal consistency value should then be the more reliable, and is in common use. I advocate, however, as a more conservative policy, the use of the larger of these two errors. On the other hand, when the ratio is not satisfactory, one must make an arbitrary assignment of weights, and with such weighting external consistency furnishes the only basis for the calculation of errors.

The writer hopes that the preceding remarks have made clear to the reader the true situation. The necessity for such a discussion may be emphasized by a few examples of the misstatements and misapplications of this matter appearing in the literature.

(1) Scarborough's article[1] is devoted to a proof of the fact that a probable error based on internal consistency varies directly with the probable errors of the original observations, whereas a probable error based on external consistency involves only ratios of these probable errors, and so is independent of their absolute magnitudes. From these facts he draws the wholly irrelevant deduction that the first method gives a correct result, but the second does not. He states, *in italics* (page 667), that $(R)_{ext}$ is simply a measure of the agreement of the points among themselves, and nothing more. After the preceding discussion it is perhaps unnecessary to remark that such agreement, or better, lack of agreement, constitutes the sole mathematical basis for the calculation of errors. The true relation between the two methods is given most simply in Eq. (12). The first method uses the agreement of the individual observations determining each point to predict what agreement will be found among the various points. The second method gives the actual agreement of the points among themselves, and is therefore an answer to the prediction. Both methods are based on the same theory, and are equally correct in principle.

(2) In Leland's text "Practical Least Squares," page 189, there is given an example of the calculation of a weighted average and of its probable error, from three observations of an angle for each of which the probable error is given. The data are

$$72^\circ\ 47'\ 43.18'' \pm 0.06''$$
$$44.01'' \pm 0.10''$$
$$43.74'' \pm 0.08''.$$

Now a mere glance at these data indicates that they show a much greater variation than is to be expected from the stated probable errors, so that unsuspected errors of some kind are undoubtedly present. Yet Leland proceeds to calculate the weighted average, and its probable error, on the basis of internal consistency. The result is $43.50'' \pm 0.04''$, and he gives this as the

final best answer. I find that $R_e/R_i=3.81$, and since $n=3$, the calculated probable error of the ratio is about 27 percent. The observed deviation from unity is thus over ten times the probable error and there is not one chance in a million of this occurring as a result of purely statistical fluctuations.

In his general discussion Leland (loc. cit. pp. 188–189) recognizes the fact that R_e may not equal R_i, but he apparently assumes that in every such case, R_i will give a more reliable result than R_e. We have seen, however, that this is not the case. On the contrary, we can form a mathematical estimate of the expected discrepancy and use this for the detection of previously unsuspected errors. In the above example one should obviously discard the stated probable errors. If, merely for convenience, the same relative weighting is retained, the correct result is $43.50'' \pm 0.16''$, as deduced from external consistency.

(3) In F. W. Clarke's monumental work on the calculation of atomic weights,[27] internal consistency is used exclusively in the calculation of probable errors and in the subsequent weighting. Now until very recently the investigation of atomic weight ratios formed a most fertile field for unsuspected constant and systematic errors. A brief examination of the data presented by Clarke is sufficient to establish this fact. I find, from sample calculations, that the ratio R_e/R_i averages about ten, so that Clarke's stated probable errors average about one-tenth of the most probable values. In certain cases such a system of analysis leads to a clearly false result for the atomic weight itself, as I have pointed out in a previous paper.[28] Thus there is occasionally an atomic weight determination by some particular investigator that is quite at variance with all other known results, but that happens to have high internal consistency. Clarke accordingly gives it a high weight, and this weight carries through to the final result, so that the investigation in question, which should have been discarded entirely, produces an appreciable change in the published final result. All of the recent reports of atomic weight committees seem to recognize the fact that the older determinations are nearly all vitiated by constant errors, and as a result the committee makes an arbitrary assignment of weights. To speak bluntly, it gives zero weight to these older determinations, regardless of their apparent probable errors. This same question arises in the following article, G.C. 1932, and in it I have adopted the same policy.

In concluding this section I give a numerical example of the calculation of errors by internal and by external consistency. For this example a situation is chosen that will incidentally bring out a point regarding weights that appeared implicitly in the preceding discussion.

Merriman[29] lists 24 observations of an angle, measured by the U. S.

[27] F. W. Clarke "A Redetermination of Atomic Weights" 4th edition (1920). Printed as Part 3, Vol. 16 of Memoirs, Nat. Acad. Science.

[28] R. T. Birge "Probable Values of the General Physical Constants," Phys. Rev. Supplement (now called Reviews of Modern Physics) **1**, 1, 1929. See pages 25 and 19–20. This article is referred to as *G.C.* 1929.

[29] M. Merriman, reference 3, page 90.

Coast-Survey, and the data need not be re-copied. The degrees and minutes also are omitted, since these are unnecessary for our purpose. The observations are of presumed equal reliability, and the arithmetic average is therefore the most probable value. This result is $49.64'' \pm 0.28''$, where the probable error has been calculated by Eq. (8).

Let us now, merely for the purpose of this illustration, divide the observations, arranged in their original order, into four sets, each comprising six observations. We may then calculate the arithmetic average of each set, and by means of the residuals,[30] a probable error for each such average. We thus get the following four *points*

$$48.82'' \pm 0.64''$$
$$48.85'' \pm 0.43''$$
$$51.19'' \pm 0.32''$$
$$49.71'' \pm 0.60''$$

Since these four sets each comprise an equal number of observations, of equal reliability, the four probable errors should be equal, except for statistical fluctuations. We may test this as follows. Treating the four errors r_i as if they were four observations of a given quantity, one obtains from Eq. (7) the probable error r' in any one of the four

$$r' = F\left(\frac{\sum(r_i - r_{\text{average}})^2}{3}\right)^{1/2}. \tag{21}$$

The proportional magnitude of r' is thus found to be 19 percent. The predicted value is given by Eq. (6) with $n = 6$, and is 19.5 percent. This very close agreement is of course mainly fortuitous, but it is certainly an indication that the various observations are of equal reliability, and that the fluctuation of the values of r_i is only statistical.

In such a situation one should disregard the variation in r_i and give each of the four points the same weight. To give them unequal weights, on the basis of the unequal values of r_i is in this case just as illogical as to give a series of observations, of presumably constant reliability, a series of weights based on the deviation of each observation from some adopted average.[31] With equal weighting for the four points, one will of course get for the final average merely the first quoted result, $49.64''$.

We now, for the purpose of the illustration, assume that the origin of the four points and of their probable errors is not known, and we proceed to give

[30] It should be noted that a residual now means the deviation of any observation from the average of its own set, *not* from the general average just quoted.

[31] These facts seem to be well recognized. See, for instance, Wright and Hayford "Adjustment of Observations" pp. 74–77, and Tuttle and Satterly, reference 4, pp. 202–203. The former text refers specifically to the Northern Boundary Commission which decided that, when a long series of observations were taken with the same instrument, under comparable conditions, the average results were to be weighted according to the number of observations composing them, and not according to the calculated probable error.

weights (P) to the points according to their probable errors, and to calculate the weighted average, A. The weight of this average is ΣP and the probable error may now be calculated in the usual two ways. By internal consistency it is

$$(R)_{int} = \frac{R_i}{(\sum P)^{1/2}} . \tag{22}$$

By external consistency it is

$$(R)_{ext} = F\left(\frac{\sum PV^2}{(N-1)\sum P}\right)^{1/2} . \tag{23}$$

The numerical results are given in the top row of Table I.

TABLE I.

A	$(R)_{int}$	$(R)_{ext}$	ratio
50.09″	0.22″	0.43″	1.95
49.64″	0.27″	0.15″	0.55
49.38″	0.21″	0.32″	1.52
Average 49.70″	0.23″	0.30″	1.34

Let us now rearrange the observations, forming the first set of six from the 1, 5, 9, etc. observations, the second from 2, 6, 10 etc. We then obtain by a similar process the values in the second row of the table. For a third division I have used observations 1, 3, 5, 7, 9, 11 for the first set, 2, 4, 6, 8, 10, 12 for the second, 13, 15, 17 etc. for the third, and 14, 16 etc. for the fourth. The results are given in the third row, and finally the average of the three methods of division in the last row. The previously quoted result obtained by the direct unweighted average ($49.64'' \pm 0.28''$) is certainly the most reliable, and the table shows that external consistency happens, in this instance, to give for the probable error a slightly more reliable average value than does internal consistency. On the other hand, the fluctuations of $(R)_{ext}$ are greater than those of $(R)_{int}$, in agreement with our previous remarks on this subject. The expected probable error in the ratio, with $N=4$, is 24 percent. From the results given in the last column of the table, the average deviation of the ratio from unity is 64 percent. There is only one chance in fourteen for such a discrepancy, and it is barely possible that all the 24 observations do not have the same reliability, in spite of our previous result to the contrary. This belief is strengthened by the fact that one of the 24 original residuals is exceptionally large, and the entire set forms a rather poor Gaussian error curve. The agreement with theory is, however, on the whole about as good as one may expect from such a small number of observations, and I have deliberately taken for this illustrative problem such a small set, since this is the situation commonly experienced in scientific work. The results show that one should be cautious in drawing any definite conclusion as to the presence of systematic errors, unless the deviation from unity of the ratio R_e/R_i is so much greater than that to be expected that there is a very small probability (say 1 in a 100 or more) that such a deviation is due to pure chance.

The Probable Error of a Function

If one has a function Z of a number of independently observed quantities z_1, z_2, z_3 etc., whose probable errors r_1, r_2, r_3 etc. are known, then the probable error R in Z is given by the well-known formula[32] for the "Propagation of Errors,"

$$R^2 = \left(\frac{\partial Z}{\partial z_1} r_1\right)^2 + \left(\frac{\partial Z}{\partial z_2} r_2\right)^2 + \left(\frac{\partial Z}{\partial z_3} r_3\right)^2 + \text{etc.} \tag{24}$$

If Z is the arithmetic mean of n observed quantities $z_1, \cdots, z_n$, and if all the probable errors r_i are equal, this equation leads to

$$R = \frac{r}{(n)^{1/2}}, \tag{25}$$

a relation used in deriving Eq. (8) from Eq. (7). If $Z = z^n$, Eq. (24) gives the well-known fact, used frequently in G.C. 1932, that the proportional error in Z is n times the proportional error in z.

An interesting application of Eq. (24), which is needed in G.C. 1932, is the following. Let us assume that we know two points, $y_1 \pm r_1$, at $x = x_1$, and $y_2 \pm r_2$, at $x = x_2$, and that we pass a straight line,

$$y = a + bx \tag{26}$$

through these two points. What are the resulting errors, r_a, r_b and R, in a, b and the function itself? To get the error in a and in b we must first express these quantities in terms of the original data. The results are

$$a = \frac{y_2 x_1 - x_2 y_1}{x_1 - x_2} \tag{27}$$

$$b = \frac{y_1 - y_2}{x_1 - x_2}. \tag{28}$$

We then apply Eq. (24), where now Z is replaced first by a and then by b, and z_1 and z_2 are replaced by y_1 and y_2. The results are

$$r_a^2 = \left(\frac{x_2 r_1}{x_2 - x_1}\right)^2 + \left(\frac{x_1 r_2}{x_1 - x_2}\right)^2 \tag{29}$$

$$r_b^2 = \left(\frac{r_1}{x_1 - x_2}\right)^2 + \left(\frac{r_2}{x_2 - x_1}\right)^2. \tag{30}$$

To obtain the error R in the function $a + bx$, we must similarly express this function directly in terms of y_1 and y_2. From Eqs. (27) and (28) we get

$$y = a + bx = y_1\left(\frac{x_2 - x}{x_2 - x_1}\right) + y_2\left(\frac{x_1 - x}{x_1 - x_2}\right). \tag{31}$$

[32] Merriman, reference 3, pp. 75–79. Palmer, reference 18, pp. 95–104.

Then, from Eq. (24), with $Z=a+bx$,

$$R^2 = \left[\frac{(x_2 - x)r_1}{x_2 - x_1}\right]^2 + \left[\frac{(x_1 - x)r_2}{x_1 - x_2}\right]^2. \tag{32}$$

When $x=x_1$, $R=r_1$, and when $x=x_2$, $R=r_2$, as should of course be the case. An actual plot of the value of R is shown in connection with curve g, Fig. 7, of G.C. 1932.

Let us now assume a more complex situation. Instead of two points, which uniquely determine a straight line, we have a number of points, through which, by means of least squares, we are to pass the *most probable* straight line. What is now the resulting error in a, in b, and in the function itself? The most general situation occurs when we have any function whatsoever, which is to be evaluated by least squares. Formulas for calculating the errors in the various constants so evaluated are given in most good texts,[33] but I have been able to find in only two places, in English, a formula for calculating the error in the function itself. In both places[34] only an entirely general formula is given, of such a character as to be rather inconvenient in actual work. I have *never* seen this formula used, nor even mentioned, although there is a distinct need for the calculation of such errors, in G.C. 1932, and I am sure in much scientific work. The following relations have therefore been obtained.

Instead of discussing the general equation, it is simpler to deduce, from elementary considerations, a specific formula for the probable error of a linear function. Let us assume that weights p have been assigned to the various points, on the basis of the given probable errors ($p=c/r^2$). Then the least squares' values of a and b, and their probable errors r_a and r_b, are given by [35]

$$a = \frac{(\sum py)(\sum px^2) - (\sum px)(\sum pxy)}{D} \tag{33}$$

$$b = \frac{(\sum p)(\sum pxy) - (\sum px)(\sum py)}{D} \tag{34}$$

where

$$D = (\sum p)(\sum px^2) - (\sum px)^2. \tag{35}$$

[33] See, for instance, Palmer, reference 18, Chap. IX, and Whittaker and Robinson, reference 3, pp. 239–242.

[34] The references are (1) Whittaker and Robinson, reference 3, near top of page 243, and (2) Wright and Hayford, reference 31 near bottom of page 138. Both published formulas give the *weight* of a function y. By Eq. (17) or (18′) one can calculate the resulting probable error, on the basis of internal or of external consistency.

[35] In G.C. 1929, page 5, the sign of both numerator and denominator in $a(=a_{01})$ and $b(=a_{11})$ was reversed, resulting in an apparent negative value of the determinant D. This may have led to confusion when D was used under the radical in Eqs. (36) and (37) ahead. Palmer, reference 18, page 112, gives the general formulas for the probable errors of the constants of any function containing only two constants.

Also

$$r_a = r_e\left(\frac{\sum px^2}{D}\right)^{1/2} \tag{36}$$

$$r_b = r_e\left(\frac{\sum p}{D}\right)^{1/2} \tag{37}$$

where

$$r_e = 0.6745\left(\frac{\sum pv^2}{(n-2)}\right)^{1/2}. \tag{38}$$

If r_a and r_b are calculated by internal consistency, r_e is replaced by $r_i(=c^{1/2})$.

From Eq. (26) it is obvious that r_a, the probable error in a, is also the error in the function y at the particular point $x=0$. We desire the error in y, at *any* point $x=\epsilon$. Let us make, in Eq. (26), the linear transformation $x'=x-\epsilon$, giving

$$y = a' + b'x'. \tag{39}$$

Then the error in y at the desired point $x'=0$ is merely the error in a'. This error, from Eq. (36), is given by

$$r_a' = r_e\left(\frac{\sum px'^2}{D}\right)^{1/2} = r_e\left(\frac{\sum p(x-\epsilon)^2}{D}\right)^{1/2} \tag{40}$$

since the determinant D is invariant to a linear transformation in x. Eq. (40) gives the probable error in any *linear* function of x, at the point $x=\epsilon$. In the case of any other rational integral function (r.i.f.) of x, one can obtain the probable error in the function at $x=\epsilon$ directly from the formula for the error in the absolute term a, by merely substituting $x'=x-\epsilon$ in place of x.[36]

The error r_a' of Eq. (40), as might be expected, is least in the vicinity of the observed points, and increases nearly linearly as one extrapolates the function in either direction. In fact the error of the function is a minimum at the "center of gravity" of the observations, defined by $\epsilon=\Sigma px/\Sigma p$, since this is the condition required to make $\Sigma p(x-\epsilon)^2$ a minimum, considering ϵ as the variable. A plot of the value of r_a' is shown, by broken lines, in each figure of G.C. 1932.

[36] The weight of the first unknown (the absolute term of a r.i.f.) is always D/A_{11}, where A_{11} is the minor determinant formed by omitting the first column and top row of the general determinant D (see Whittaker and Robinson, reference 3, pp. 231 and 241). For a r.i.f. of degree f, D is given by

$$\begin{vmatrix} \sum p & \sum px & \sum px^2 \cdots & \sum px^f \\ \sum px & \sum px^2 & \cdots\cdots\cdots & \\ \vdots & & & \vdots \\ \sum px^f & \cdots\cdots\cdots\cdots\cdots & & \sum px^{2f} \end{vmatrix}$$

I have checked these results by showing that in the case of the r.i.f. of zero, first and second degree, the general formula mentioned in footnote 34 actually reduces to the explicit form D/A_{11}, with x' substituted for x in A_{11}.

In scientific work it is quite customary to evaluate the coefficients of a function by least squares' methods. Rather more rarely, the probable error in each coefficient is similarly evaluated. The accuracy of the resulting function, on the other hand, is usually judged roughly from the deviations of the data (the residuals) from the calculated function. The equations just quoted furnish a quantitative evaluation of this accuracy, not only in the vicinity of the data, but also in the extrapolated region on either side of the data. It seems to the writer that such information must necessarily be of value in much scientific work.

PROBABLE VALUES OF e, h, e/m AND α

By Raymond T. Birge
University of California, Berkeley

(Received February 18, 1932)

Abstract

Calculation of general physical constants.—The calculation of the general physical constants offers a rich and almost untouched field of investigation. In general each equation for the calculation of a given constant contains also other general constants, and therefore does not, strictly speaking, evaluate any one of them. In 1929 the writer solved this problem by a method of successive approximations, i.e., the constant first evaluated was that one least dependent on others. This constant then became an *auxiliary* constant, of known magnitude, in the equations for other constants. The method breaks down when the probable error of any auxiliary constant is not small compared to the experimental error. This is the case in the calculation of h, which always involves e and in which the error in h is usually due *mainly* to the error in e. The present paper gives in detail a logically correct method for the *simultaneous* evaluation of e and h, from the several known functional relations between these two constants. The method was suggested first by W. N. Bond, but was not carried by him to its logical conclusion. The writer also disagrees with portions of Bond's calculations and with the final values of the constants that Bond adopts.

General theoretical relation between e and h.—Every so-called method for the evaluation of h yields an equation of the type $h=A_n e^n$ (1), in which A_n is an experimentally determined magnitude, and $n=3/3$, $4/3$, or $5/3$ according to the method employed. For least squares' calculations it is necessary to use an equation *linear* in the unknowns, and in the present case the most convenient form for such an equation is obtained by the introduction of a new parameter h_n, defined by $h_n=A_n e_0^n$, where e_0 is a tentatively adopted value of e. One then gets the linear equation $h_n=h-(h_0\,\Delta e/e_0)n$. The intercept at $n=0$ gives h, and from the slope one gets e $(=\Delta e+e_0)$. In the present work we assume $e_0=4.770\times10^{-10}$ *es* units, $h_0=6.547\times10^{-27}$ erg. sec. Plotted h_n:n curves, with the probable error in the *function* h_n indicated by broken lines, are shown in Figs. 1–7 and 10. The value of e so obtained is entirely *independent* of the adopted e_0 and we thus have a new method for the evaluation of e, of greater potential accuracy than any direct measurement. To obtain final most probable values of e and h, our original Eq. (1) is transformed to $e=a_m h^m$ (2), in which $m=1/n$ and $a_m=(A_n)^{-1/n}$. Then $m=0$ corresponds to a direct determination of e and one can thus include in a single equation *all* data on e and h, and can determine final most probable values. This is the essential extension of Bond's method.

Least squares procedure.—The necessary least squares' procedure involves most of the questions discussed in the article preceding this, on the calculation of errors by least squares. In order to make several corrections in Bond's work, and to indicate the proper procedure, a sample calculation is carried through in detail, using Bond's adopted data. The values of h and e so obtained agree with Bond's results, but the probable errors are quite different from those deduced by him. This solution is shown in Fig. 1, curve *a*. The h_n point for $n=5/3$ depends almost entirely upon the value adopted for e/m. The points resulting from $e/m=1.761\times10^7$ and $e/m=1.769$ are shown on each figure. From $h_{6/3}$ one can calculate the fine structure constant α, and the value of $h_{6/3}$ corresponding to Eddington's theoretical value $1/\alpha=137$ is shown in each figure. Bond uses the results of 36 investigations and these are shown by circles in Fig. 1. The three resulting points (arithmetic averages of observations) for the h_n:n curve are shown by crosses.

Additional solutions of Bond's data.—Bond weights each point according to the number of observations composing it. When the reliability of the observations also is included in the weighting, one obtains curve *b*, Fig. 2. The values of e, h, e/m, and $1/\alpha$, resulting from each solution, are collected in Table II (page 242). The first two solutions differ mainly in respect to the probable errors. Three new investigations on e/m have recently appeared, and Fig. 3 gives a solution corresponding exactly to curve *b*, except for the inclusion of these three new observations. The new values of e/m are all "low" and the calculated value of $1/\alpha$ (137.115 ± 0.054), for curve *c*, Fig. 3, deviates appreciably from Eddington's value.

Solutions based on 1929 data, using Eq. (1).—The writer believes that values of constants should be derived mainly, if not wholly, from recent experimental work. The data used by Bond include practically all work, new and old, with each investigation given equal weight. The writer in 1929 calculated a most probable value of h_n for each of the six available methods. Three of these (ionization potentials, photoelectric effect and c_2) correspond to $n=3/3$, two (x-rays and σ) to $n=4/3$, and one (Rydberg constant) to $n=5/3$. The six adopted values of h_n are first given equal weight and combined by the new method. The results are shown in curve *d*, Fig. 4. Then the six observations are weighted according to their probable errors, and the results appear in curve *e*, Fig. 5. The point $h_{5/3}$ is based on $e/m=1.761\times10^7$, since this value now seems well established. The chief difference between curves *d* and *e* and the preceding is due to the adoption of this value of e/m, the "spectroscopic" value of the 1929 work.

Solutions based on present data, using Eq. (1).—The data now available for each of the six methods are critically examined, and new most probable values of each h_n obtained. If, as has been customary during the past two years, one assumes that the correct wave-lengths of x-rays are those determined from grating measurements, the x-ray point is moved from $n=4/3$ to $n=3/3$, and an accompanying necessary correction gives a very high resulting value of $h_{3/3}$. At the same time there remains at $n=4/3$ only the very "low" $h_{4/3}$ observation derived from σ. Due mainly to these facts, the resulting solution, curve *f*, Fig. 6, exhibits a definite incompatibility of the data ($r_e/r_i=2.14$). The results therefore cannot be used, and to get curve *g*, Fig. 7, the two discordant methods (x-rays and σ) are discarded. The resulting value of e (4.7732 ± 0.0072) represents the best *independent* determination by the new method, on the assumption that the x-ray method belongs at $n=3/3$. Fig. 7 also includes a curve based on $1/\alpha=137$. This curve requires $e/m=1.7679$, and leads to $e=4.7824$, $h=6.5670$.

Solutions based on Eq. (2).—By introducing a new parameter $e_m=a_m h_0^m$, one can obtain the linear equation $e_m=e-(e_0\Delta h/h_0)m$, which is to be used in the least squares' solution of Eq. (2). e_m:m curves are shown in Figs. 8, 9, and 11. A simple relation is derived, giving the value of e_m corresponding to any h_n. Curve *h*, Fig. 8, results from the three h_n values used for curve *g* (converted to e_m values), *plus* the oil-drop value $e_d=4.768\pm0.005$. The resulting constants are practically identical with those adopted in 1929, as given in Table II. Curve *i*, Fig. 9, is based on the 1929 data, and represents the constants that would have been obtained in 1929, had the data been properly handled. There is a considerable change in some of the probable errors, but not in the constants.

New solutions with x-ray method at $n=4/3$.—Very recent observations by Bearden, as well as the discrepancies noted in curve *f*, give convincing evidence that the crystal wave-lengths of x-rays are correct, rather than the grating values. On this assumption the x-ray method belongs at $n=4/3$ and a recalculation of the data of curve *f*, with this one change, gives curve *j*, Fig. 10, which exhibits no inconsistency of data. The resulting value of e (4.7688 ± 0.0059) is the most reliable resulting from the new method, and is in remarkable agreement with the oil-drop value. Curve *k*, Fig. 11, represents the e_m:m curve corresponding to curve *j*, with the oil-drop value added.

General conclusions.—(1) The values of e, h and $1/\alpha$ depend in a very direct way upon the value adopted for e/m. In the past ten years all investigations on e/m, with

one exception, have given results in the close vicinity of 1.761, and with this value one obtains (curve *k*) $1/\alpha = 137.307 \pm 0.048$. Conversely, the assumption $1/\alpha = 137$ practically requires $e/m = 1.768$, for which there is now very little experimental evidence. This is the most important result of the present investigation. (2) It would appear that the crystal wave-lengths of x-rays are correct, and the close agreement of the resulting value of e (curve *j*) with the oil-drop value is strong evidence that there is no unsuspected error of any significance in Millikan's work. (3) Curve *k* gives the most probable values of the constants ($h = 6.5443 \pm 0.0091$, $e = 4.7688 \pm 0.0040$, $1/\alpha = 137.307 \pm 0.048$ and $e/m = 1.7611 \pm 0.0009$), but these differ so little from the values adopted in 1929 that *no change is advocated at this time.* However, the probable errors of all constants involving h^x/e^y, as published in 1929, are more or less in error, and a general equation is given for the correct calculation of such probable errors. A full report on all the general constants, including derived constants, will be published by a committee of the National Research Council, whenever this seems necessary.

Introduction

THE calculation of probable values of the various general physical constants offers a rich and hitherto almost untouched field of investigation. Much time and effort have been devoted to the experimental evaluation of each constant, and as a result there is a really extensive amount of experimental material available to the computer. How this material should be handled, mathematically, in order to obtain the most reliable values of the desired constants, constitutes a problem of considerable complexity. This is quite aside from the question of the proper selection and weighting of the data, which involves primarily the individual judgment of the computer. If each general constant could be evaluated by means of an equation that contained no other general constant, the situation would be comparatively simple. Unfortunately this is not the case. Each such equation normally contains one or more additional constants, and in a previous investigation on this subject[1] I have called these "auxiliary constants."

In this previous work I attempted to solve the problem arising from the presence of these auxiliary constants by calculating first the constant that depends least on the other constants, and then consistently adopting this value in all further calculations. In most cases the probable error in the constant being evaluated is due chiefly to the direct experimental errors, and depends only in a minor degree on the errors of the auxiliary constants. In such a situation the procedure adopted in 1929 constitutes essentially a "method of successive approximations" for the simultaneous evaluation of all the constants, and is theoretically quite correct. The method, however, breaks down completely when the error in the constant being investigated is due *chiefly* to the error in one or more auxiliary constants. This is just the situation in the case of the Planck constant h. There is no known method for evaluating h that does not involve the electronic charge e as an auxiliary constant, and the resulting error in h is, in most cases, due mainly to the error in e. The writer pointed out this situation, in the 1929 work, but failed to recognize the procedure necessary to give the best results.

[1] R. T. Birge "Probable Values of the General Physical Constants," Phys. Rev. Supplement (now called Reviews of Modern Physics) **1**, 1, 1929. This article will be denoted as G.C. 1929.

A general statement of the correct procedure is as follows. If there exists a theoretical equation containing, let us say, e, h, and e/m, this equation cannot be said to furnish a value of any one of these constants. If, however, there exist three different theoretical relations involving these three constants, one can use the three relations simultaneously for the evaluation of all three constants. If more than three such relations exist, the *most probable* values of the constants are best determined by least squares. It has been remarked that the mathematically correct way to determine the atomic weights of the elements is to make one simultaneous least squares' evaluation of all atomic weights, using all measured mass ratios. The practical objection is that the necessary calculations might easily take centuries to carry out.

The situation, as it exists in connection with the fundamental physical constants, is far less complex, due to the smaller number of such constants. In fact it appears now as though e, h, and e/m are the only important constants that should be simultaneously evaluated in a straight-forward manner. In the present paper the problem is still further simplified by assuming a value of e/m, so that only e and h remain as unknowns. The chief purpose of the paper is to present what appears to be a theoretically correct method for finding the most probable values of these two constants. It is hoped that this general method will be used in future calculations, and the method is therefore given in detail. The paper also includes a re-examination of available data, and a new calculation of probable values of e, h, and the fine structure constant α. The final conclusion of the paper is that there is at the moment no sufficient cause for advocating a change in the writer's 1929 values of these constants, but there is certainly cause for discarding the very crude method used in these 1929 calculations.

The major principle underlying the new method of calculation has been suggested by Bond,[2,3] but in his two articles Bond fails to carry the theory to its logical conclusion. The writer does not agree at all with Bond's choice of data, nor with his weighting of the data, and it is for that reason that a re-examination of available material has been made. All of Bond's calculations are carried out by least squares' methods, but in several cases he seems to have used formulas not applicable to the situation. In fact there is scarcely an English text on the subject of least squares that even mentions some of the formulas needed for these calculations. It is partly for this reason that the paper just preceding this[4] has been written. The work of the present paper really constitutes an interesting example of the application of least squares, and I open the paper with a discussion of the fundamental equations, and the correct procedure for handling them, using Bond's own data as illustration. This is followed by a recalculation of the data adopted in my 1929 article, and finally by a discussion and calculation of what appears to be the best data available at present.

[2] W. N. Bond, Phil. Mag. **10**, 994 (1930).

[3] W. N. Bond, Phil. Mag. **12**, 632 (1931).

[4] R. T. Birge, "The Calculation of Errors by the Method of Least Squares." To be referred to as L.S. 1932.

The General Theoretical Relation Between e and h

Each so-called method for the evaluation of h yields an equation of the type

$$h = A_n e^n \tag{1}$$

where e is the electronic charge and A_n is an experimentally determined magnitude. The value of n, for each method, is listed in the last column of the table on page 57 of *G. C.* 1929. One observes that three methods (ionization potentials, photoelectric effect and c_2) involve $n = 3/3$, two methods (x-rays and σ) involve $n = 4/3$, and one method (Rydberg constant) involves $n = 5/3$. Now, as noted in the Introduction, Eq. (1) is, strictly speaking, not an equation for the evaluation of h. It does, however, permit the simultaneous evaluation of e and h, since at least two values of n, together with the corresponding values of A_n, are available. This is the major point brought out in Bond's papers. His procedure, however, does not lead to the final most probable values of e and h because he omits all directly measured values of e, such as for example that obtained by the oil-drop method. What Bond does obtain is a new value of e, entirely *independent* of the oil-drop value.

To get the final most probable values of e and h, Eq. (1) should be rewritten as follows[5]

$$e = a_m h^m \tag{2}$$

where

$$m = 1/n \text{ and } a_m = (A_n)^{-1/n}. \tag{3}$$

Then $m = 0$ corresponds to any *directly* determined value of e, while $m = 3/5$, $3/4$ and $3/3$ correspond to the above-mentioned values $n = 5/3$, $4/3$ and $3/3$. It is thus possible to include simultaneously *all* work on e and h, and so to obtain final most probable values of both constants. There is, however, an important advantage in the preliminary use of Eq. (1). By means of this equation one can determine a value of e that should be consistent with the directly observed value. Only when such consistency exists is it permissible to use Eq. (2) to obtain a final most probable value of e. For that reason a number of calculations will be made, using Eq. (1), and various deductions will be drawn from the results. Then a selected number of these calculations will be repeated, using Eq. (2).

Eqs (1) and (2) are non-linear in the unknowns, e and h, and the proper method of procedure, in order to obtain least squares' values of such unknowns, is to be found in every good text. The probable errors in the resulting values of e and h are given by formulas that appear in many, but not all, texts on least squares. Finally, in order to evaluate the fine structure constant α, it is necessary to know the value of a certain $f(n)$ at the point $n = 2$, and the probable error in α follows directly from the probable error of the *function* at this point. It is therefore necessary to determine the probable

[5] The writer is indebted to Professor R. B. Brode for this important suggestion.

error of a function whose coefficients (e and h) have been simultaneously evaluated by least squares. This type of error, as noted in L. S. 1932, is scarcely mentioned in the literature, and has accordingly been discussed rather fully in that article.

The Least Squares Procedure

As already noted, Eq. (1) involves h and e as unknown quantites, and n and A_n as known quantities. When an equation is non-linear in the unknowns, the standard least squares procedure[6] is to adopt approximately correct values of the unknowns, and then to calculate the proper corrections to be applied to these tentative values. Using a Taylor's expansion one obtains the necessary linear observational equations. In this case let e_0 and h_0 be the tentatively adopted values. Also let

$$\left.\begin{aligned} e &= e_0 + \Delta e \\ h &= h_0 + \Delta h \end{aligned}\right\} \tag{4}$$

Eq. (1) is now to be written as

$$f(h, e) = he^{-n} = A_n. \tag{5}$$

The general form of the resulting observational equation is

$$\left(\frac{\partial f}{\partial h}\right)_0 \Delta h + \left(\frac{\partial f}{\partial e}\right)_0 \Delta e = l \tag{6}$$

where

$$l = f(h, e) - f(h_0, e_0) = A_n - h_0 e_0^{-n}. \tag{7}$$

Eqs. (5), (6) and (7) give

$$e_0^{-n}\Delta h - nh_0 e_0^{-(n+1)}\Delta e = A_n - h_0 e_0^{-n}. \tag{8}$$

Eq. (8) is not in a convenient form for use as an observational equation, for the following reason. The various precision methods for determining h involve *proportional* errors in A_n (and in h) of the same order of magnitude. The absolute value of A_n (and accordingly its absolute error) varies, however, as e^{-n}, as shown by Eq. (5). Now for each increase of 1/3 in the value of n, e^{-n} and A_n increase about 1280 fold. Hence the *weight* of each observational equation (which is to be taken inversely proportional to the square of the absolute probable error) varies enormously with n, and this introduces a troublesome feature in the numerical calculations.

The whole situation is greatly simplified by the introduction of a new parameter

$$h_n \equiv A_n e_0^n. \tag{9}$$

Substituting in Eq. (8) the value of A_n from Eq. (9) and using Eq. (4) we obtain

$$h_n = h - \left(h_0 \frac{\Delta e}{e_0}\right) n. \tag{10}$$

[6] G. C. Comstock "Method of Least Squares" pp. 21–23. Merriman "Method of Least Squares" pp. 200–204. W. W. Johnson "Theory of Errors etc.," pp. 93–94.

Eq. (10) is of the form

$$y = a + bx \tag{11}$$

where

$$\left.\begin{array}{ll} a = h & b = -\left(h_0 \dfrac{\Delta e}{e_0}\right) \\ y = h_n & x = n \end{array}\right\} \tag{12}$$

This new observational equation is linear in the unknowns, a and b, and the absolute value of y, or h_n, changes only very slightly with n. Hence the absolute error in h_n is always of the same order of magnitude, and the various observational equations have comparable weights. Moreover, as discussed in footnote 20 of L. S. 1932, the least squares' formulas are strictly applicable, since the errors of observation are confined to the ordinate y.

In terms of a graph, the various observations h_n obtained from Eq. (9) are to be plotted as ordinates against n as abscissa. The best straight line through the data is then to be calculated by least squares. The intercept of this line on the h_n axis, at $n=0$, gives the desired value of h. From the slope of the line we can obtain Δe and so, by Eq. (4), the desired value of e. This is the method used by Bond.

In my 1929 paper, h was calculated by the use of Eq. (9). That is, a most probable value of e (here called e_0) was assumed, and then from each observed A_n one calculated an h_n. The weighted average of the various h_n values was taken as the most probable value of h. Such a procedure is equivalent to fitting the h_n values to the best *horizontal* straight line (h_n = constant). It is legitimate only when h_n, plotted against n, shows entirely irregular variations (due solely to the experimental errors in A_n). If now the assumed value of e_0 is *not* correct, the values of h_n will show a regular trend with n. The least squares' calculation of the best straight line through the various points is essentially the calculation of a new value of e_0 (to be called e) which, used in Eq. (9), will give a new set of values of h_n showing the least possible trend with n, and is, simultaneously, the calculation of the best average value (to be called h) of this new set of values of h_n. Even if the trend of h_n with n is so small as to be undetected on an h_n:n plot, the least squares' solution should be carried out, since only in this way can the most reliable values of e and h be obtained.[7]

In G. C. 1929 six methods for calculating h were discussed. These methods, as already noted, correspond to values of $n = 3/3$, $4/3$ and $5/3$. There is a seventh method, corresponding to $n = 6/3$, which was not mentioned because its precision was not yet comparable with the others. This seventh method concerns the fine structure constant α, which is given by the equation

[7] The reader is reminded that in such a method the value e_0 first assumed is naturally some directly determined value of e, but the final calculated value of e quite ignores this assumed value. Hence, as emphasized by Bond, the procedure now under discussion constitutes an entirely *new and independent* method for evaluating e, but as noted earlier, a really proper procedure for determining a final most probable value of e must use *both* the new value and the directly determined value. This is accomplished by the use of Eq. (2), in place of Eq. (1).

$$\alpha = 2\pi e^2/hc \tag{13}$$

Hence

$$h = \left(\frac{2\pi}{\alpha c}\right)e^2. \tag{14}$$

The most accurate direct measurement of α is that by Paschen,[8] who found $\Delta\nu_{He} = R_{He}\cdot\alpha^2/16 = 0.3645 \pm 0.0045$ cm^{-1}. Using Paschen's value[9] $R_{He} = 109722.14 \pm 0.04$ cm^{-1}, one obtains $\alpha = (7.291 \pm 0.045)\times 10^{-3}$. It is more convenient to state α in terms of its reciprocal, and in this case $1/\alpha = 137.16 \pm 0.85$. With $e = 4.770\times 10^{-10}$, this yields $h = (6.54 \pm 0.04)\times 10^{-27}$ erg·sec., exclusive of the error in e. This is a much less accurate value of h than that obtained by the other six methods, as will appear in the discussion to follow. Accordingly this seventh method is omitted, and we shall use our final calculated values of h and e to determine a most probable value of $1/\alpha$. The most important conclusion of Bond's two articles is that the available data are in agreement with Eddington's theory[10] that $1/\alpha = 137$. The value of $1/\alpha$ will accordingly be discussed with special reference to this point.

In terms of Eq. (9) we may write Eq. (14) as

$$h_2 = \left(\frac{2\pi}{\alpha c}\right)e_0^2. \tag{15}$$

Hence, after evaluating the constants of Eq. (10), we use that equation to calculate h_2 and then Eq. (15) to calculate $1/\alpha$. The probable error in h_2 is the error in the function

$$f(n) \equiv h - \left(h_0\frac{\Delta e}{e_0}\right)n \tag{16}$$

at the point $n = 2$, and from Eq. (15) the proportional error in $1/\alpha$, aside from the negligibly small error in c, equals the proportional error in h_2. The results published by Bond indicate that he used an incorrect formula for the error in $1/\alpha$, as well as for the errors in other quantities. It therefore seems advisable to go through all steps of a typical solution. For this purpose I shall use the 36 values of h_n quoted on page 634 of Bond,[3] without any comment as to the scientific value of the data.[11] Such comment will appear later.

The data listed by Bond consist of 14 observations of h_n corresponding to $n = 3/3$, 9 corresponding to $n = 4/3$, and 13 corresponding to $n = 5/3$. These observations are plotted as circles in Fig. 1 of the present article. Double and triple circles indicate two and three coincident observations. Now we desire a solution of Eq. (10), and in this case the final data consist of *three* values of h_n (to be called *points*), one for each of the three values of n (3/3, 4/3 and

[8] F. Paschen, Ann. d. Physik **50**, 901 (1916). See also **82**, 689 (1927).

[9] See G.C. 1929, pp. 45 and 46.

[10] A. S. Eddington, Proc. Roy. Soc. **A126**, 696 (1930).

[11] In calculating values of h_n by means of Eq. (9), Bond uses $e_0 = 4.770\times 10^{-10}$ *es* units, and $h_0 = 6.547\times 10^{-27}$ erg·sec. These are the values given in G.C. 1929, and Bond's paper, like the present one, is concerned with possible changes in these values.

5/3). The function obviously cannot have two or more values for a single value of n, and any observed scattering of observations at one value of n must be due solely to experimental errors. On the other hand, a *variation* of h_n with n may well exist, due to the fact that our adopted e_0 is not the best value of e, and it is this variation that we are trying to evaluate. The first step is then to calculate a weighted average value of h_n for each different value of n. Bond gives each original observation[12] unit weight, so that each resulting point is to be given a weight (p) equal merely to the number of observations of which it is the arithmetic average. These points[13] are denoted by crosses in Fig. 1, and are listed in Table I, col. 2.

TABLE I.

1 n	2 h_n(obs) curve a	3 p	4 h_n(calc)	5 r_a'	6 r curve b	7 p	8 h_n(calc)	9 r_a'
0			6.55751	0.00960			6.55681	0.00635
3/3	6.5473	14	6.54561	0.00307	0.0019	2.77	6.54630	0.00243
4/3	6.5364	9	6.54165	0.00204	0.0034	0.86	6.54280	0.00132
5/3	6.5395	13	6.53768	0.00317	0.00085	13.84	6.53930	0.00112
6/3			6.53373	0.00521			6.53579	0.00210

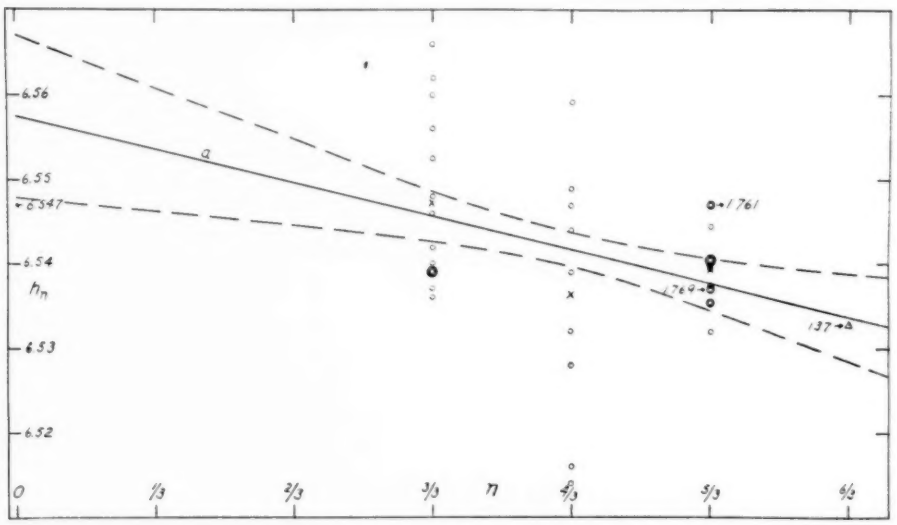

Fig. 1. Solution of Eq. (17). 36 observations (circles) used by Bond, with each resulting point (cross) weighted according to the number of observations composing it. $h=6.5575$, $e=4.7787$. Probable error of the calculated *function* h_n shown by broken lines.

In discussing least squares' formulas, it is more convenient to use Eq. (10) in the general form of Eq. (11), to which it is connected by the relations given

[12] Just as in L.S. 1932, I shall attempt to avoid confusion by using the word *observation* for each h_n, as determined directly from one investigation, and the word *point* for the average value of h_n from all observations at any one value of n.

[13] In discussing h and e, the factors 10^{-27} and 10^{-10} respectively are omitted whenever this can cause no misunderstanding.

in Eq. (12). The least squares' formulas for a and b, and for their probable errors r_a and r_b, are given as Eqs. (33) to (38) of L.S. 1932. With each point given the *arbitrary* weight shown in Table I, col. 3, it is possible to calculate all probable errors only on the basis of external consistency[14] i.e., by the use of r_e, as defined by Eq. (38), L.S. 1932. Eq. (10), thus evaluated, is plotted as curve a in Fig. 1, and is given by

$$h_n = 6.5575 \pm 0.0096 - (0.0119 \pm 0.0071)n. \tag{17}$$

Hence $h_0(\Delta e/e_0) = +0.0119 \pm 0.0071$, and since $h_0 = 6.547$ and $e_0 = 4.770$, we find $\Delta e = +0.0087 \pm 0.0052$ or $e = 4.7787 \pm 0.0052$. Also, from Eq. (17), $h = 6.5575 \pm 0.0096$.

In place of these results, Bond gives $h = 6.5575 \pm 0.0053$, and $e = 4.7787 \pm 0.0029$. His published errors seem to have been obtained in the following way. Instead of getting the three average values of h_n (the points), he uses Eqs. (33) and (34), L.S. 1932, as though there were 36 different points. It is easily shown that this procedure will give values of a and b identical with those obtained by the method of weighting used in Table I, col. 3. The calculated errors will however *not* be the same, as the results show. This is due to the fact that his value of r_e is not correct. The rest of Eqs. (36) and (37), L.S. 1932, is unchanged. Bond calculates r_e from an equation of the correct form, i.e.,

$$r_e = 0.6745\left(\frac{\sum pv^2}{(s-2)}\right)^{1/2} \tag{18}$$

but he uses $p = 1$, $s = 36$,[15] and $v =$ the deviation of *each* original observation (36 in all) from its value calculated by Eq. (17). In the correct interpretation of Eq. (18), (or of Eq. (38), L.S. 1932), v is the deviation of each of the three *points* (average values of h_n) from its value calculated by Eq. (17), p is the weight of the point, and $s = 3$. By using the former incorrect interpretation of Eq. (18), I have been able to reproduce all Bond's published probable errors, and therefore deduce that that was the procedure he used. That his procedure cannot be correct is immediately evident from the fact that if all observed h_n values correspond to the same value of n (giving only one point), no solution at all would be possible. An increase in the number of observations at any one value of n is useful only in giving a more reliable point. The accuracy of the resulting coefficients in Eq. (10) depends solely on the number of *different* values of n represented (i.e., on the number of points), and the accuracy with which these points fit a linear relation. These facts are brought out more clearly in the actual calculations given later.

The values of h_n, as calculated from Eq. (17), are given in col. 4 of Table I, and from the calculated h_2 we obtain $1/\alpha$ by writing Eq. (15) as

$$1/\alpha = \left(\frac{c}{2\pi e_0^2}\right)h_2 = \frac{h_2}{47.68575 \times 10^{-30}}. \tag{19}$$

[14] See section on "Internal versus External Consistency," L.S. 1932.

[15] Throughout this article s is used for the number of observations, (or the number of points), in place of the usual n, in order to avoid confusion with n in Eqs. (1) and (10).

The result is $1/\alpha = 137.016$. The proportional probable error in $1/\alpha$, as has been pointed out in connection with Eq. (15), equals the proportional error in the function h_n at $n=2$. The formula for obtaining such an error of a function is given by Eq. (40) L.S. 1932. It is interesting to calculate the probable error in the function h_n, not only at $n=0$ and 2, but also at each value of n considered. This error is given in col. 5, Table I, and is indicated by the broken lines of Fig. 1. The error has been calculated and is similarly shown in each of the succeeding figures. In the case of Eq. (17), $h_2 = (6.5337 \pm 0.0052) \times 10^{-27}$. Hence, by Eq. (19), $1/\alpha = 137.01_6 \pm 0.10_9$ (Bond gives $137.01_7 \pm 0.05_9$).

It is of interest to calculate also the value of e/m that corresponds to $h_{5/3}$, as given by Eq. (10), From the formula for the Rydberg constant[16] and from Eq. (9) we obtain

$$e/m = \frac{2\pi^2 \cdot c_0^5}{R_\infty c^2 h_{5/3}^3} = \frac{4.94214 \times 10^{-72}}{h_{5/3}^3}. \tag{20}$$

Conversely, in obtaining *observed* values of $h_{5/3}$ from any observed value of e/m, we have the relation

$$h_{5/3} = \frac{7.90625 \times 10^{-27}}{(e/m \times 10^{-7})^{1/3}}. \tag{21}$$

In taking the cube root, it is convenient to use the form $e/m \times 10^{-7} = 1.761$ or thereabouts, and the factor in the numerator of Eq. (21) corresponds to this.

The proportional error in the calculated value of e/m is, from Eq. (20), three times the proportional error in the calculated $h_{5/3}$. Using this value of $h_{5/3}$ and its error, as given in Table I, we obtain from Eq. (20), $e/m = (1.7686 \pm 0.0026) \times 10^7$. This is merely the value of e/m that it is necessary to assume, in Bohr's formula for the Rydberg constant, in order to be consistent with the values of h and e calculated from Eq. (17). As in the case also of $1/\alpha$, the writer made such a calculation of e/m in his 1929 work, but it was not evident at that time how the corresponding error in $1/\alpha$ and in e/m should be calculated.

It will appear from the discussion to follow that the most probable values of h and e depend chiefly on the value adopted for e/m. In 1929 the writer gave *two* values, the so-called "spectroscopic" value 1.761×10^7 *em* units, and the "deflection" value 1.769×10^7. These, by Eq. (21), correspond to values $h_{5/3} = 6.54714 \times 10^{-27}$ and 6.53724×10^{-27} respectively. These two points are located by arrows on each figure of this article. The 1929 value, $h = 6.547$ (our present h_0), is also shown on each figure. Finally, by Eq. (19), Eddington's value $1/\alpha = 137$ corresponds to $h_2 = 6.53295 \times 10^{-27}$, and this point is indicated by a triangle, on each figure. The discrepancy between each new value of e, and the previously adopted $e_0 = 4.770$ can be judged from the slope of the h_n curve. A horizontal line would indicate $e = e_0$.

This completes the discussion of the various formulas needed for the solution of Eq. (1), and we pass now to a consideration of the scientific aspects of the problem.

[16] G.C. 1929, page 49.

Other Possible Values of e and h Based on Bond's Observational Data

Bond[3] has chosen, from all available data, the 36 values listed in his article and used as illustrative material in the preceding section. Even assuming that these 36 observations should be chosen, it does not follow by any means that the values of e and h corresponding to Fig. 1 are the most probable. In the first place, Bond gives each observation unit weight and, as has been discussed, this is equivalent to weighting each of the three resulting points merely according to the number of observations that it represents. It is however quite obvious that the reliability of any point h_n, for a given value of n, should be judged by the *consistency* of the observations as well as by their number. Accordingly we shall first consider other possible ways in which these 36 observations, and the resulting 3 points, may be weighted, and shall draw certain conclusions as to the most desirable method of weighting. These conclusions will then be applied to the newer data now available.

The most rigorous method of weighting should start with the 36 observations. Each should be weighted according to what is judged to be its probable error. Such a probable error should in turn be based not only on the purely accidental errors of observation, but also on all other possible errors, constant or systematic, so far as these can be estimated. Since the 36 observations listed by Bond include practically all work in the field, old as well as new, such an evaluation would be very difficult to make in an intelligent manner, and does not seem worth the effort. The method will however be applied to the solutions given later.

(1) Curve *b*. A simpler method of weighting the observations is that adopted by Bond. Each observation is given equal weight and the arithmetic average of all observations at a given value of n is calculated. We thus obtain again the three points listed in col. 2 of Table I. We shall, however, weight each of these *points* according to its probable error r, and not merely according to the number of observations included by it.[17] The values of r calculated from Eq. (13) L.S. 1932, and the resulting values of $p(=c/r^2)$, are listed in cols. 6 and 7 of Table I. In this, as in succeeding calculations, the arbitrary constant c is taken as 10^{-5}. A comparison of cols. 3 and 7 of Table I shows how greatly the weighting is modified when the consistency of the observations is considered.

The least squares' solution of Eq. (10) is now

$$h_n = 6.5568 \pm 0.0063 - (0.0105 \pm 0.0041)n. \qquad (22)$$

This equation is plotted as curve *b*, Fig. 2, together with the probable error of the function, at each value of n. The arrow attached to each observed point measures its assumed error as listed in Table I, col. 6. The calculated values of h_n and the probable errors of these calculated values are listed in cols. 8 and 9 of Table I. All of these probable errors associated with Eq. (22) have been calculated, just as in the case of Eq. (17), on the basis of external

[17] See Eqs. (13), (14), (15), L.S. 1932 and accompanying discussion.

consistency. In the present case, however, the three individual points have been given weights based on their probable errors and it is accordingly possible to calculate errors also on the basis of internal consistency.

The subject of internal versus external consistency has been discussed fully in L.S. 1932, and the reader is reminded that there is a constant ratio between all errors, as calculated by the two systems. This ratio is computed most simply by using a hypothetical point of unit weight. For this point the probable error on the basis of external consistency is given by r_e, Eq. (18), and the corresponding error on the basis of internal consistency is, in this article, always given by

$$r_i = (c)^{1/2} = (10^{-5})^{1/2} = 3.162 \times 10^{-3}. \tag{23}$$

In the case of the solution now under discussion, the data needed for Eq. (18) are given in cols. 2, 7 and 8 of Table I, with $s=3$. The result is $r_e=4.188 \times 10^{-3}$. Hence $r_e/r_i=1.32$.

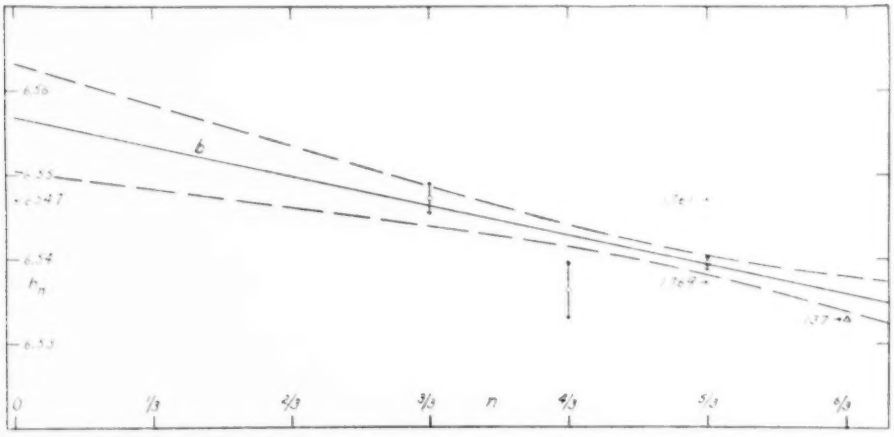

Fig. 2. Solution of Eq. (22). Same data as in Fig. 1, but points weighted according to the consistency of the observations, as well as their number. Probable error of each point shown by attached arrows. $h=6.5568$, $e=4.7777$.

Now, as shown in L.S. 1932, these two methods of calculating errors should agree, except for statistical fluctuations, i.e., the ratio r_e/r_i should be unity. Any considerable deviation from unity is an almost sure indication of the presence of constant or systematic errors in the data. In the present case the 32 percent deviation from unity can well be attributed to chance, since the proportional probable error in the ratio is given by $0.4769/s^{1/2}=27.5$ percent, with $s=3$.[18] We therefore conclude that there is here no definite evidence of systematic errors. In such a situation I adopt the policy suggested in

[18] This formula for the probable error of the ratio appears as Eq. (6), L.S. 1932. As discussed in footnote 26 of that article, the probable error in any one of the actual constants, such as h or e, is more uncertain than this, since the uncertainty in the weight of the constant must also be included. Bond[3] does not include this second source of uncertainty and uses the equation just quoted with $s=36$. He thus obtains an apparent reliability for his stated probable errors greatly in excess of the true value.

L.S. 1932, and use r_e, since it is larger than r_i. The errors given in Eq. (22) are therefore retained, and the various results that may be derived from this equation are as follows,

$$h = 6.5568 \pm 0.0063 \quad 1/\alpha = 137.060 \pm 0.044$$
$$e = 4.7777 \pm 0.0030 \quad e/m = 1.7673 \pm 0.0009.$$

These two solutions differ mainly in the values of the probable errors and illustrate in a striking way how dependent such results are upon the adopted method of weighting.

For convenience of comparison, the values of the four constants, for all the different solutions discussed in this paper, are listed in Table II. Each of these solutions is plotted, and the curve designation appears in the first column of the table. The second column gives the number of the equation stating each solution.

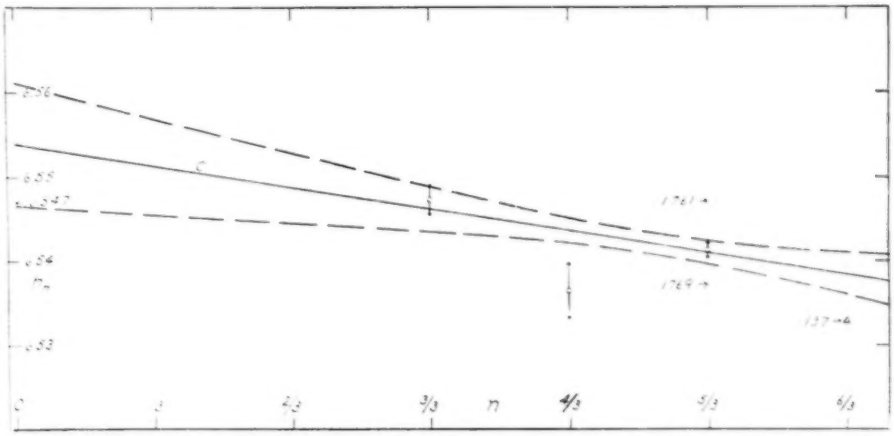

Fig. 3. Solution of Eq. (24). Same data and method of weighting as in Fig. 2. but with the addition of three new observations of $h_{5/3}$ (from values of e/m). $h=6.5539$, $e=4.7756$.

(2) Curve c. In listing the 36 observations used in the preceding analysis, Bond apparently attempted to include all available data. Even since the writing of his paper, however, three new investigations on e/m have been completed. Campbell and Houston[19] from the Zeeman effect, get 1.7577; Perry and Chaffee,[20] from electrostatic acceleration of electrons get 1.761 ± 0.001; and Kirchner,[21] by the same method, gets 1.7598 ± 0.0025. It is of interest to add these three new observations, and to obtain a new solution by the same procedure (including method of weighting) as used for curve *b*. The points at $n=3/3$ and $4/3$ and their weighting are unchanged. For $n=5/3$ we have $h_n=6.5413 \pm 0.00096$, giving $p=10.85$. The solution is

$$h_n = 6.5539 \pm 0.0072 - (0.0077 \pm 0.0047)n. \qquad (24)$$

[19] J. S. Campbell and W. V. Houston, Phys. Rev. **38,** 581 (1931). Their final value (Phys. Rev. **39,** 601 (1932)) is 1.7579 ± 0.0025, but this small change does not affect the final average value of $h_{5/3}$.

[20] C. T. Perry and E. L. Chaffee, Phys. Rev. **36,** 904 (1930).

[21] F. Kirchner, Ann. d. Physik **(5) 8,** 975 (1931).

This is shown by curve *c*, Fig. 3, and the various resulting constants are tabulated in Table II. The errors are again calculated on the basis of external consistency, since the ratio r_e/r_i is now 1.49.

TABLE II.

Curve	Eq.	h	e	$1/\alpha$	e/m
a	17	6.5575 ± 0.0096	4.7787 ± 0.0052	137.016 ± 0.109	1.7686 ± 0.0026
b	22	6.5568 ± 0.0063	4.7777 ± 0.0030	137.060 ± 0.044	1.7673 ± 0.0009
c	24	6.5539 ± 0.0072	4.7756 ± 0.0034	137.115 ± 0.054	1.7659 ± 0.0011
d	25	6.5568 ± 0.0064	4.7753 ± 0.0037	137.193 ± 0.087	1.7631 ± 0.0021
e	26	6.5360 ± 0.0117	4.7652 ± 0.0053	137.338 ± 0.064	1.7612 ± 0.0009
f	36	6.563 ± 0.022	4.777 ± 0.010	137.22 ± 0.12	1.7611 ± 0.0021
g	44	6.5543 ± 0.0166	4.7732 ± 0.0072	137.266 ± 0.079	1.761 ± 0.001
h	55	6.5464 ± 0.0095	4.7696 ± 0.0041	137.302 ± 0.048	1.7610 ± 0.0009
1929		6.547 ± 0.008	4.770 ± 0.005	137.29 ± 0.11	1.761 ± 0.001
i	56	6.5441 ± 0.0079	4.7688 ± 0.0035	137.303 ± 0.046	1.7612 ± 0.0009
j	57	6.5443 ± 0.0133	4.7688 ± 0.0059	137.305 ± 0.069	1.7611 ± 0.0010
k	58	6.5443 ± 0.0091	4.7688 ± 0.0040	137.307 ± 0.048	1.7611 ± 0.0009

Certain conclusions may be drawn from the constants for this solution. In the first place, the value of $1/\alpha$ differs from 137 by more than twice the probable error, so that the data can hardly be said to support Eddington's theory. The new calculated value of e agrees with the directly determined value (4.770 ± 0.005) as well as is to be expected, considering the probable errors. This is very important, for it indicates that there are no serious systematic errors in the direct measurements. Points like these will, however, be discussed more critically after various other solutions have been obtained.

SOLUTIONS BASED ON 1929 DATA, USING EQ. (1)

The results given by curve *c* might be considered the best that can be obtained, if it is wise to include all possible investigations and to give to each investigation the same weight. The writer, however, emphatically disagrees with such a choice of data. At the risk of repeating remarks made in previous articles, the following statement of policy is presented.

Most important constants have been measured many times, and in some cases by a number of different methods. The only object of repeating previous work is to obtain a greater precision in the result. Very often this precision is attained by the elimination of various sources of constant and systematic error that there is reason to believe existed in the earlier work. Now the assignment of equal weights to the various results is merely the assumption of equal reliability. If, however, the newer results are no more reliable than the older, it would appear that these newer investigations represent more or less wasted effort. The computer may remark that he does not wish to pass judgment on the various investigations, and therefore gives them all equal weight. I feel, however, that the computer is practically compelled to pass such judgment; otherwise the computation should not be made.

When one thus proceeds to consider the available data, it becomes immediately evident that the newer investigations are in general entitled to far greater weight than the older. The original experimental work on a given

constant is often very brilliant, but the numerical results are likely to be marred by the presence of constant errors that is revealed only by later and more detailed investigation. For this reason it is, I believe, generally agreed at the present time that values of constants should be based almost wholly on the most recent work in the field. This is the theory adopted for my 1929 work, and it is the theory on which the rest of this article is based.

(1) Curve *d*. In the case of curves *a*, *b* and *c*, all observations corresponding to a given value of n have been averaged together. It has, however, been noted that three different methods for evaluating h correspond to $n=3/3$, two to $n=4/3$, and one to $n=5/3$. It seems best to consider separately the value of h_n resulting from each of these six methods. That procedure was adopted in 1929 and the resulting values are listed in the table on page 57 of G.C. 1929. The weighted average obtained then was based on the assumption that there is

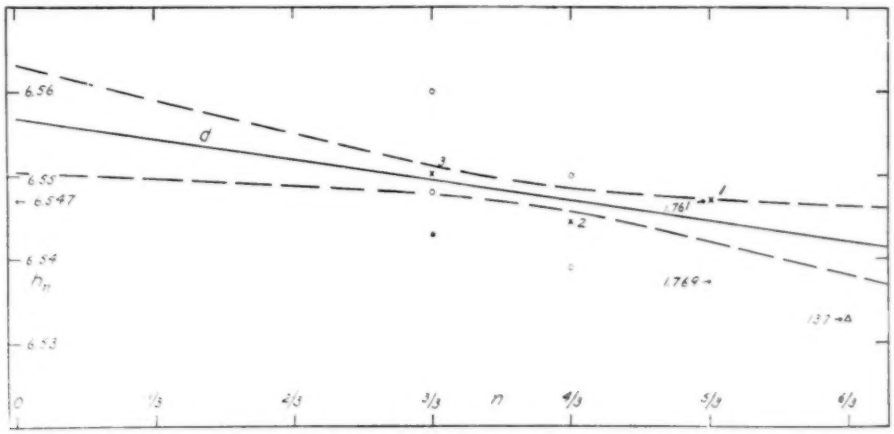

Fig. 4. Solution of Eq. (25). Most probable value of h_n, from each of the six methods, as adopted in 1929. Each point weighted according to the number of observations composing it. $h=6.5568$, $e=4.7753$.

no variation of h_n with n. It is therefore of interest to consider what values of h and e would have been obtained in 1929, had the data been handled by the new method. In making this calculation, all six values of h_n are first given the same weight, and are later weighted according to their probable errors. This is done deliberately, to show the change in the resulting values of e, h etc., brought about by the change of weighting.

Proceeding to the first calculation, we obtain the final value of $h_{3/3}$ from the arithmetic average of the three observations (methods) for which $n=3/3$. This point is given a weight of three. Similarly the final value of $h_{4/3}$ is given a weight of two, and $h_{5/3}$ a weight of unity. It is impossible here to base the relative weights of these three points on the consistency of the observations, since in the case of $h_{5/3}$ there is only one observation. The three points and their weights are listed in Table III, cols. 2 and 3. The solution of Eq. (10) is now given by

$$h_n = 6.5568 \pm 0.0064 - (0.0073 \pm 0.0051)n. \qquad (25)$$

This is plotted as curve *d*, Fig. 4, and the resulting constants are listed in Table II. In Fig. 4 the six observations are indicated by circles, and the three resulting points by crosses. The divergence of this solution from the preceding is due chiefly to the fact that the value of $h_{5/3}$ is now based solely on the "spectroscopic" value of e/m ($=1.761$). As is evident from Fig. 4, this raises the value of $1/\alpha$ considerably above the 137 figure.

TABLE III.

1 n	2 h_n(obs) curve *d*	3 p	4 h_n(obs) curve *e*	5 p	6 h_n (obs) curve *f*	7 p
3/3	6.5503	3	6.5472 ± 0.0058	0.30	6.5605 ± 0.0046	0.47
4/3	6.5445	2	6.5409 ± 0.0037	0.73	6.539 ± 0.0040	0.625
5/3	6.5471	1	6.5471 ± 0.0012	6.94	6.5471 ± 0.0012	6.94

(2) Curve *e*. We now proceed to consider the probable error in each of the six observations, in order to obtain a more reliable basis for weighting. The error we desire is that in h_n. As shown by Eq. (9) this is merely the error in h due to all sources except the electronic charge. It is *not* the error listed in the table, page 57 of *G. C.* 1929, since there the error in e is included. From the discussion on pp. 48–57 of G.C. 1929, the desired errors are as follows.

(a) In the case of the Rydberg constant, the errors in c and R_∞ are negligible. The proportional error in $h_{5/3}$ is one-third the proportional error in e/m. Assuming $e/m = 1.761 \pm 0.001$, we obtain the value of $h_{5/3}$ listed in Table III, col. 4. The corresponding weight p is listed in col. 5.

(b) At $n = 4/3$ there are two methods. The first involves the Stefan-Boltzmann constant σ, for which 5.735 ± 0.011 was adopted in 1929. The only significant proportional error in $h_{4/3}$ is one-third the proportional error in σ. Hence[22] from σ, we obtain $h_{4/3} = 6.539 \pm 0.00405$ ($p = 0.610$). The second method involves measurements on the continuous spectrum of x-rays. The value adopted in 1929 is 6.550 ± 0.009 ($p = 0.124$). This result was obtained by giving an arbitrary weight of two to the value 6.559, as calculated by the writer from the experimental results of Duane, Palmer and Yeh, and a weight of unity to the similarly calculated value 6.532, based on Wagner's work. The error ± 0.009 just quoted is, as noted on page 53 of G.C. 1929, the regular least squares' probable error, and is therefore the error desired here.[23] The weighted average result of these two methods is 6.5409. The ratio r_e/r_i is 0.75 and an error based on internal consistency (r_i) is accordingly used. The final result for $h_{4/3}$ is given in Table III, col. 4. The weight of this result, since r_i is used, is of course merely the sum of the weights of the two methods.

(c) There are three available methods for obtaining the point at $n = 3/3$. The first is that of ionization potentials, and the only significant error is that

[22] It was found later that the correct error is ± 0.00419. This corrected value has been used in curve *j* ahead, but it did not seem necessary to make a recalculation here. See also footnotes 37 and 43.

[23] In the 1929 work I accidentally neglected to compound this error with that due to e. The value that should have been used in 1929 is ± 0.011.

in the measured voltage. The resulting value of h is 6.560 ± 0.0131 ($p = 0.0583$). In the case of the photoelectric effect the various assumed errors are discussed on pp. 53–54 of G.C. 1929. The result is 6.543 ± 0.0073 ($p = 0.1876$). In the third method the only error that concerns us is that due to c_2 itself. With $c_2 = 1.432 \pm 0.003$, one obtains $h = 6.548 \pm 0.0137$ ($p = 0.0533$). The weighted average of these three values of $h_{3/3}$ is 6.5472 and the ratio $r_e/r_i = 0.54$. Here again the results of the various methods, for a given value of n, are more consistent than is to be expected from theory, although the deviation is not unreasonable. Therefore we again base the final probable error on r_i rather than on r_e. The resulting value of $h_{3/3}$ is given in Table III, col. 4, and as before the weight is merely the sum of the three weights just given.

Using the data listed in Table III, cols. 4 and 5, we obtain

$$h_n = 6.5360 \pm 0.0117 + (0.00653 \pm 0.00725)n. \qquad (26)$$

In this case $r_e/r_i = 0.89$ and the quoted errors are based on r_i. This equation is plotted as curve e in Fig. 5. The adopted probable errors of each of the

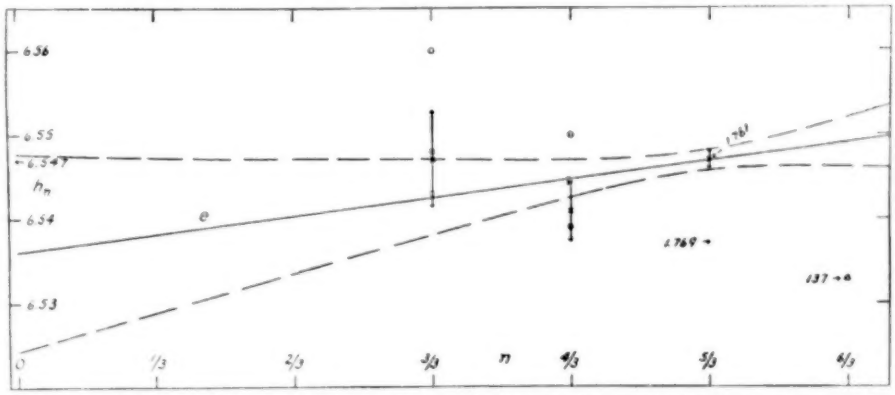

Fig. 5. Solution of Eq. (26). Same data as in Fig. 4, but probable error of each of the six observations computed, and all weights based on probable errors. $h = 6.5360$, $e = 4.7652$.

three points are indicated, but not those of the six observations (circles). It is very gratifying to find that r_e/r_i is less than unity, in the case of the observations and also in the case of the points. The results thus show no indication of constant or systematic errors in any one of the six adopted values of h_n.

The resulting values of the various constants are given in Table II. Since they do not include any direct measurement of e, they cannot be called the results that should have been obtained in 1929. (Such results are given in curve i ahead.) The only fact that seems worthy of mention is the rather large difference in the values of the constants and also of their probable errors, in curves d and e. This difference is due solely to the change from arbitrary equal weighting for all six methods to a more logical system of weighting.

Solutions Based on Present Data, Using Eq. (1)

We proceed now to a re-examination of all available data, beginning with the methods for which $n = 3/3$.

(a) Ionization potentials. There seems to be no new precision measurements since 1929, comparable in accuracy with that by Lawrence,[24] and I therefore again adopt his determination of the ionization potential of mercury, leading as before to

$$h_{3/3}(\text{ion. pot.}) = 6.560 \pm 0.0131 \ (p = 0.0583). \tag{27}$$

(b) Photoelectric effect. In the 1929 work, the only investigation considered was that by Lukirsky and Prilezaev. Their value of $h_{3/3}$, as already quoted, is 6.543 ± 0.0073, exclusive of the error in e. As shown in the discussion on pp. 53–54 of G.C. 1929, it was difficult, due to lack of information, to assign an error in this case, and it seems best now to raise the error slightly to ± 0.010. (This was the 1929 error, *including* the error in e.) Since 1929 there has appeared a new investigation by Olpin.[25] His published result, based on a graphical solution of the voltage: frequency curve, is 6.541. Dr. Olpin has kindly sent me his original data for this curve, and I find for the least squares' solution, $h_{3/3} = 6.561 \pm 0.029$, and this result will be used in the following.

The original precision work on h, from the photoelectric effect, is that by Millikan.[26] A considerable amount of experimental data is given in his article, with several different results. It has seemed desirable to recalculate these results, using least squares' methods for the calculations and the weighting. There is no indication of systematic error in the several results, and my final weighted average of all the data is 6.560 ± 0.037. The actual calculations are rather extensive and it is unnecessary to present them at this time.

For the final value of $h_{3/3}$, as determined from the photoelectric effect, I use the weighted average of these three investigations. This result is

$$h_{3/3}(\text{photoelectric}) = 6.546 \pm 0.0092 \ (p = 0.1181). \tag{28}$$

The ratio r_e/r_i equals 0.337, thus showing a considerably greater consistency of the three individual results than is to be expected on the average. The quoted error is based on r_i.

(c) The radiation constant c_2. There is no new work on this important constant, and I accordingly retain the 1929 value 1.432 ± 0.006, which yields

$$h_{3/3}(c_2) = 6.548 \pm 0.0137 \ (p = 0.0533) \tag{29}$$

as previously quoted.

(d) X-rays. In 1929, the value of h resulting from observations on the continuous spectrum of x-rays was listed under $n = 4/3$. The defining equation in this case is[27]

$$h_n = \frac{e_0 pq V' \lambda 10^8}{c^2}. \tag{30}$$

[24] E. O. Lawrence, Phys. Rev. **28**, 947 (1926).
[25] A. R. Olpin, Phys. Rev. **36**, 251 (1930), see p. 284.
[26] R. A. Millikan, Phys. Rev. **7**, 355 (1916).
[27] Compare Eqs. (11) and (12), p. 51 of G.C. 1929.

Now, as noted in the 1929 work,[28] there are two methods for determining λ. One is by means of the Bragg law

$$\lambda = 2d \sin \theta \tag{31}$$

in which the calculated grating space d involves e_0 to the one-third power, so that h_n varies with e_0 raised to the 4/3 power. This is the method[29] used in 1929, and the x-ray result was accordingly listed under $n = 4/3$.

The second method for determining the λ of Eq. (30) is by means of ruled gratings. This is a direct determination, quite independent of any assumed value of e, so that h_n now varies as $e^{3/3}$ and should be listed under $n = 3/3$. The quantity actually observed in the x-ray experiments now under discussion was sin θ of Eq. (31), so that it was possible to calculate only a Bragg law wave-length. We may, however, assume that there exists for all wave-lengths the proportional discrepancy found by various investigators for certain lines, as measured with a ruled grating and with a calcite crystal. The two most accurate investigations are those by Bearden[30] and by Cork.[31] The former found the grating λ of two lines to be respectively 0.23 percent and 0.24 percent higher than the crystal values. The latter found, for different lines, discrepancies of 0.288 and 0.305 percent. If we now assume, in agreement with prevailing opinion, that wave-lengths as measured by ruled gratings are actually correct, then all values of h_n, as calculated from x-ray continuous spectra are to be raised by 0.265 percent,—the unweighted average of the above four results. In terms of 10^{-27} erg. sec, the correction is +0.01740.

In addition to the x-ray work actually used in 1929, we now have available the result of Feder.[32] This investigation is mentioned in footnote 16[a], page 53 of G. C. 1929, but appeared too late to use there. Since the work is a repetition of that by Wagner, it seems reasonable to use it as a substitute for Wagner's result. Feder's value of h, using my 1929 values of the auxiliary constants, is 6.5463. His work appears to be of accuracy comparable to that of Duane, Palmer and Yeh, and I accordingly give these two investigations equal weight and adopt the arithmetic average $h = 6.55265$. Since both results are based on crystal wave-lengths, we must now add 0.01740, thus getting

$$h_{3/3}(x\text{-rays}) = 6.5700 \pm 0.0063 \ (p = 0.252). \tag{32}$$

The probable error is calculated directly from the agreement of the two results (i.e., on external consistency), just as was done in 1929.

[28] See pp. 52 and 39–43 of G.C. 1929.

[29] See Eq. (13) p. 52 of G.C. 1929. The discussion following Eq. (16) page 52 is not very logical, again due to a failure to realize that what was being calculated was $h_{4/3}$ and not h itself. Since 4.770 was used in both Eqs. (13) and (15) for what should have been called e_0, the resulting value of h necessarily varied as $e_0^{4/3}$ and should have been denoted $h_{4/3}$.

[30] J. A. Bearden, Proc. Nat. Acad. Sci. **15,** 528 (1929).

[31] J. M. Cork, Phys. Rev. **35,** 1456 (1930).

[32] H. Feder, Ann. d. Physik **(5) 1,** 497 (1929).

This completes the discussion of the four methods available for determining $h_{3/3}$. The weighted average of the results listed in Eqs. (27) (28) (29) and (32) is

$$h_{3/3} = 6.5605 \pm 0.0046 \quad (p = 0.47). \tag{33}$$

The ratio $r_e/r_i = 0.92$, so that the four results are self-consistent, in spite of the suspiciously high x-ray value. We shall, however, find immediately that this shift of the x-ray result from $n = 4/3$ to $n = 3/3$ brings about a definite discrepancy in the final collected results. Later in this paper we shall return again to a consideration of the proper value of λ to use in Eq. (30) and we shall see that it is fairly probable that the x-ray method should be left at $n = 4/3$.

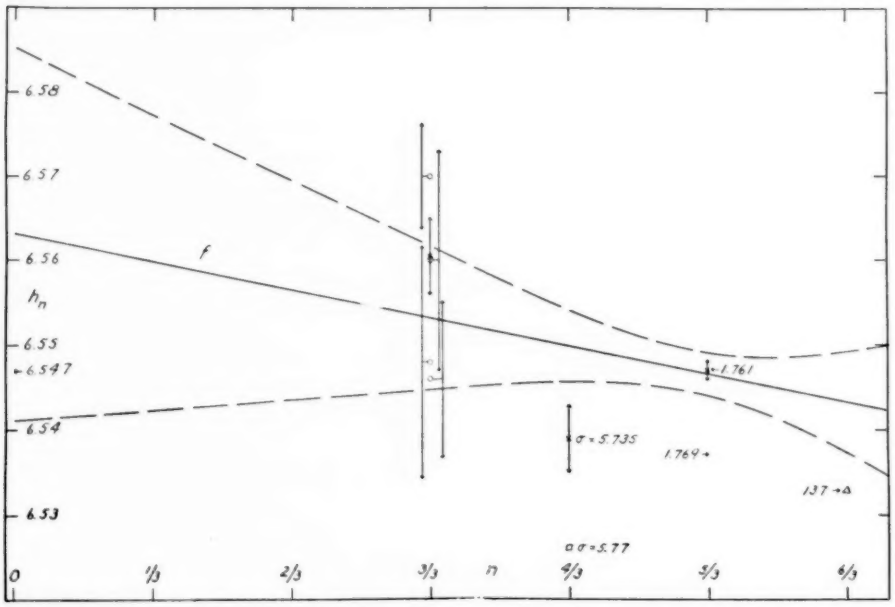

Fig. 6. Solution of Eq. (36). Six observations resulting from a re-examination of all available data. Probable errors of observations and of resulting points shown by arrows. X-ray observation at $n = 3/3$, in place of former $n = 4/3$. $h = 6.563$, $e = 4.777$, but data inconsistent.

Assuming for the present that there are four methods for obtaining h that belong under $n = 3/3$, there remains only one method at $n = 4/3$, and one at $n = 5/3$. The one method at $n = 4/3$ is based on the total radiation constant σ. There is still considerable uncertainty regarding the best value of this important constant. The reader is referred to pp. 55–57 of G. C. 1929 for a discussion of the situation at that time. As noted on page 57, Hoare's value of 5.735, which was published too late to use in 1929, but which fortunately agreed exactly with the assumed value, appears to be more reliable than any previous determination. Since then Mendenhall[33] has obtained 5.79 (error not stated) and C. Müller, in a general article[34] on this constant,

[33] C. E. Mendenhall, Phys. Rev. **34**, 502 (1929).

[34] Wien-Harms Handbuch d. Exp. Physik **9**, 427–455 (1929).

lists a recent value of his own, 5.77 ± 0.03, of which I have seen no further reference. He adopts 5.77 as the most probable value, in agreement with Ladenburg's earlier estimate.[35] A value of 5.77 corresponds to $h_{4/3} = 6.5256$, and this point is plotted on Fig. 6. It obviously is inconsistent with the results from all other methods. Even $\sigma = 5.735$ seems too high, and the indirect value[36] of σ, as calculated in 1929, is 5.7139×10^{-5}. As discussed then, the chief experimental uncertainty seems to lie in the corrections to be applied for incomplete absorption of the receiver. Hoare claims that his method eliminates such corrections, and it appears to the writer, as it did in 1929, that 5.735 is still the most reliable experimental value.[36a] The probable error is, under the circumstances, very uncertain, and merely for convenience I shall retain the 1929 value. Hence[37]

$$h_{4/3} = 6.539 \pm 0.0040 \quad (p = 0.625). \tag{34}$$

In the case of $n = 5/3$, the only known method for obtaining h is from the Rydberg constant formula. The error in $h_{5/3}$ is due almost solely to the error in e/m, and as shown in curve *e*, based on the 1929 data, this point is, apparently, so accurately known that the final values of e and h depend primarily upon the adopted value $e/m = 1.761 \pm 0.001$. The three new determinations of e/m, since 1929, have been listed in connection with curve *c*. They are all 1.761 or lower. Two of them are obtained from the acceleration of free electrons in an electric field, and constitute the first "low" values of e/m obtained by a non-spectroscopic method. It is still extremely important to obtain a really reliable value of e/m from magnetic deflection. In the meantime the presumption of evidence is that there is only one value[38] of e/m, and that this is the so-called "spectroscopic" value. It is quite possible that this value is 1.760, or even lower, but it has seemed better to retain, for the present, the 1929 value, 1.761 ± 0.001. Hence

$$h_{5/3} = 6.5471 \pm 0.0012 \quad (p = 6.94) \tag{35}$$

as used in curve *e*.

[35] Geiger and Scheel, Handbuch d. Physik **23**, 305 (1926). See p. 56 of G.C. 1929.

[36] See Table *b*, p. 61, G.C. 1929. With the values of the principal constants, including e, adopted in 1929, one has $\sigma = 16.03418 \times 10^{-84}/h_{4/3}^3$ and $h_{4/3} = 2.521635 \times 10^{-28}/\sigma^{1/3}$.

[36a] In a very recent investigation, Phil. Mag. **13**, 380 (1932) Hoare gets 5.737, from 50 extremely consistent determinations.

[37] The writer must apologize for several small but annoying discrepancies in this article. Many solutions have been obtained that, for lack of any significant features, are not published, and the total calculations are quite extensive. As a result, there have been several instances where brief calculations were unintentionally repeated, with slightly different results due to the employment of a different number of significant figures. Thus, in a previous section, Eq. (34) was quoted as ± 0.00405 ($p = 0.610$), and as stated in footnote 22, it happens that both these results are slightly in error.

[38] Several writers have implicitly criticized the assumption of two different values of e/m, as made in G.C. 1929. I think every one admits that there can be but one *correct* value, and the sole reason for adopting the two values, in 1929, was to call attention, as emphatically as possible, to the obvious discrepancy in the experimental results. The extensive discussion and investigations of e/m that have since appeared in the literature constitute sufficient evidence that this object has been attained.

(1) Curve *f*. This completes the discussion of present available data. The three resulting points are given in Eqs. (33) (34) and (35), and are listed in cols. 6 and 7 of Table III. The solution of these data is

$$h_n = 6.563 \pm 0.022 - (0.010 \pm 0.014)n. \tag{36}$$

This is plotted as curve *f*, Fig. 6, and the resulting constants are given in Table II. The value of the ratio r_e/r_i is 2.14, a deviation from unity of 4.14 times the probable error. There is only one chance in 190 of such an excess, and we thus have here clear evidence of an incompatibility of the three h_n points. A glance at curve *f* shows that this unwelcome result is due jointly to the abnormally low value of $h_{4/3}$, and the abnormally high value of that one of the four observations composing $h_{3/3}$ deduced from the x-ray method. It is difficult to explain the low value of $h_{4/3}$, and it has just been noted that most reviewers favor $\sigma = 5.77$, which leads to the even lower value $h_{4/3} = 6.5256$. On the other hand, the high value of $h_{3/3}$ from the x-ray method is due mainly to the correction computed from the assumption that the true x-ray wave-lengths are those given by ruled gratings. The writer believes that the situation shown by curve *f* constitutes definite independent evidence against such an assumption.

(2) Curve *g*. The next logical step, under the circumstances, seems to be to ignore entirely both the σ and the x-ray result. This leaves three observations at $n = 3/3$, given by Eqs. (27) (28) and (29). The new weighted average is

$$h_{3/3} = 6.5500 \pm 0.0066 \ (p = 0.23). \tag{37}$$

The ratio $r_e/r_i = 0.425$, as contrasted with 0.92 when four methods were included.

The one remaining point is $h_{5/3}$, as given by Eq. (35). With only two points and two undetermined constants we now abandon least squares' methods and make a direct calculation of the intercept and slope of the $h_n : n$ curve. These constants, and their errors, are given by Eqs. (27) to (30) of L. S. 1932, and the error in the function itself by Eq. (32). In the present case the $y_1 \pm r_1$ of L. S. 1932 may be denoted $h_3 \pm r_3$, where h_3 stands for $h_{3/3}$, and $y_2 \pm r_2$ may be denoted $h_5 \pm r_5$, where h_5 stands for $h_{5/3}$. Also $x_1 = 3/3$ and $x_2 = 5/3$. Hence Eqs. (27) to (32), *L. S.* 1929, become respectively

$$a = 2.5h_3 - 1.5h_5 \tag{38}$$

$$b = 1.5(h_5 - h_3) \tag{39}$$

$$r_a = [(2.5r_3)^2 + (1.5r_5)^2]^{1/2} \tag{40}$$

$$r_b = [(1.5r_5)^2 + (1.5r_3)^2]^{1/2} \tag{41}$$

$$y = h_n = h_3(2.5 - 1.5n) + h_5(1.5n - 1.5) \tag{42}$$

$$r_n = ([(2.5 - 1.5n)r_3]^2 + [(1.5n - 1.5)r_5]^2)^{1/2}. \tag{43}$$

With $h_3 \pm r_3$ given by Eq. (37) and $h_5 \pm r_5$ by Eq. (35) one gets

$$y = h_n = 6.5543 \pm 0.0166 - (0.00435 \pm 0.0100)n. \tag{44}$$

This equation is plotted as curve g in Fig. 7, and the errors in the function, calculated by Eq. (43), are shown as usual by broken lines. This error of the function must of course equal the error of any observed point, at the corresponding value of n. Hence in Fig. 7 the broken lines touch the arrows indicating the assumed errors r_3 and r_5. The magnitude of $1/\alpha$ and its error are calculated in the usual way. In the case of e/m, however, we have only the value originally assumed, i.e., 1.761 ± 0.001. The various constants and their errors are listed in Table II.

If the grating values of x-ray lines are correct, with the result that the value of h_n belongs at $n = 3/3$ and is with justification omitted, then curve g represents the best independent determination of e. This curve gives also the clearest illustration of the dependence of e and h upon the value adopted

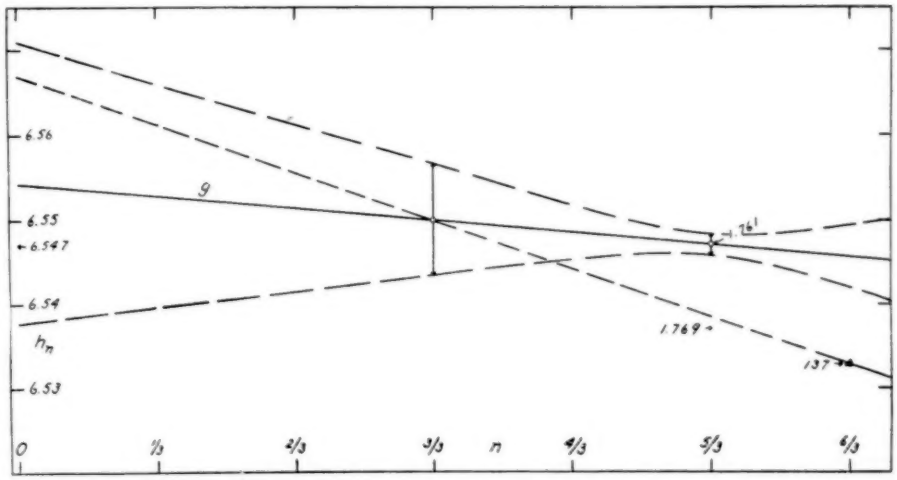

Fig. 7. Solution of Eq. (44). Same data as Fig. 6, but with x-ray and σ observations discarded. $h = 6.5543$, $e = 4.7732$. Broken straight line represents necessary solution ($h = 6.5670$, $e = 4.7824$) if $1/\alpha = 137$.

for e/m. Taking $h_{3/3} = 6.5500$ as one fixed point, we may for instance assume $h_{3/3} = 6.53295$ (corresponding to $1/\alpha = 137$) for the other. The resulting curve is indicated by broken lines in Fig. 7. Its equation is

$$h_n = 6.56705 - 0.01705n \tag{45}$$

and the resulting constants are $e/m = 1.7679$, $e = 4.7824$, $h = 6.5670$.

It thus seems highly probable that if Eddington's theory concerning $1/\alpha$ is true, the correct value of e/m must be approximately 1.768, and as already discussed, there is now very little experimental evidence in favor of such a value. If, however, future work should establish the correctness of Eddington's theory, then it follows almost inevitably that the correct value of e is about 0.26 percent higher than our previously adopted $e_0 = 4.770$, and that h is about 0.30 percent higher than $h_0 = 6.547$.

Returning once more to the assumptions made in deriving curve g, we note that the probable errors in the resulting constants are greater than in

the case of most of the previous solutions. This is due primarily to the fact that the curve is based on two points only, instead of on three. The value of e (4.7732 ± 0.0072) agrees, however, with the oil-drop value (4.768 ± 0.005) well within the probable errors, and we have therefore achieved our main objective, i.e., we have obtained a new and entirely independent value of e and have shown that it is compatible with the previously accepted value. Hence we may now proceed to the calculation of "most probable" values of e and h, by the use of Eq. (2) which allows the inclusion of a directly measured value of e.

Solutions Based on Eq. (2)

Eq. (2) is of the same form as Eq. (1), with h and e merely interchanged. Hence the least squares' treatment of Eq. (2) leads to a linear equation like Eq. (10), with e and h interchanged, i.e.,

$$e_m = e - \left(e_0 \frac{\Delta h}{h_0}\right) m \tag{46}$$

where e_m, in analogy with Eq. (9), is a new parameter defined by

$$e_m \equiv a_m {h_0}^m. \tag{47}$$

Also, as stated in Eq. (3), $m = 1/n$, and $a_m = (A_n)^{-1/n}$.

Ordinarily one would evaluate e_m by means of the value of a_m that can be calculated from the experimental data. But in this case we know all the needed values of h_n and from them it is possible to get, by a simple calculation, the corresponding values of e_m. Thus, from Eqs. (9) (47) and (3) we obtain

$$A_n = \frac{h_n}{{e_0}^n} = \frac{h_0}{{e_m}^n}. \tag{48}$$

Hence

$$\frac{e_m}{e_0} = \left(\frac{h_n}{h_0}\right)^{-1/n}. \tag{49}$$

Then, writing $e_m = e_0 + \delta e$, and $h_n = h_0 + \delta h$, where δe and δh are small quantities, Eq. (49) becomes

$$\frac{\delta e}{e_0} = -\frac{1}{n}\frac{\delta h}{h_0}. \tag{50}$$

Putting in the numerical values of e_0 and h_0 (4.770 and 6.547), we get

$$\delta e = -\frac{0.72857\,\delta h}{n} \tag{51}$$

where

$$\delta h = h_n - h_0 \text{ and } e_m = e_0 + \delta e. \tag{52}$$

This gives the value of e_m corresponding to any assumed value of h_n. Also, in Eq. (51), if δh represents the probable error in h_n, then δe represents the corresponding probable error in e_m.

As in the case of the $h_n{:}n$ curve, it is convenient to obtain explicit formulas for the calculation of $1/\alpha$ and e/m from the $e_m{:}m$ curve. By analogy with Eqs. (14) (15) and (19) we write

$$1/\alpha = \left(\frac{ch_0}{2\pi}\right)\frac{1}{e_{1/2}{}^2} = \frac{3.12385 \times 10^{-17}}{e_{1/2}{}^2} \tag{53}$$

and the proportional error in $1/\alpha$ is therefore twice the proportional error in $e_{1/2}$, which in turn is the error in the function e_m at $m=1/2$. For $1/\alpha=137$, we have $e_{1/2}=4.77512\times10^{-10}$, and this point is shown on the plotted e_m curves.

Similarly, e/m may be calculated from the value of e_m at $m=3/5$. Corresponding to Eq. (20) we write

$$e/m = \frac{2\pi^2 e_{3/5}{}^5}{R_\infty c^2 h_0{}^3} = (7.13175 \times 10^{53})e_{3/5}{}^5 \tag{54}$$

and the proportional error in e/m is five times that in $e_{3/5}$. If $e/m=1.761\times10^7$, $e_{3/5}=4.76996\times10^{-10}$, and if $e/m=1.769\times10^7$, $e_{3/5}=4.77427\times10^{-10}$. These two points also are shown on the plotted e_m curves.

(1) Curve *h*. For the first solution of Eq. (46) I use the data of curve *g*, *plus* the best directly measured value of e (to be called e_d). This latter is the oil-drop value[39] 4.768 ± 0.005, and not the 4.770 value adopted in 1929, for the following reason. This latter value is a weighted average of the oil-drop result, 4.768, and a value calculated from the grating measurements of x-ray wave-lengths. The more recent work of Bearden and of Cork, as already noted, indicates a discrepancy of 0.265 percent in the crystal and grating wave-lengths. The discrepancy in the resulting value of e is three times as great, or 0.8 percent. It is generally agreed at the present time, and the calculations in this article support the opinion, that the observed discrepancy in the x-ray wave-lengths cannot be due to an error of 0.8 percent in the adopted value of e (4.770), but must be due to some other cause. In that situation it is not possible to calculate a value of e from grating wave-lengths, and the only directly determined value that is available is the oil-drop result 4.768.

The needed data are, accordingly,

$$e_{3/3} = 4.7678 \pm 0.0048\ (p = 0.43),\ \text{from}\ h_{3/3} = 6.5500 \pm 0.0066,$$
$$e_{3/5} = 4.76996 \pm 0.00052\ (p = 37.0),\ \text{from}\ h_{5/3} = 6.5471 \pm 0.0012,$$
$$e_d = 4.768 \pm 0.005\ (p = 0.40).$$

This last point is of course to be plotted at $m=0$, since it is independent of the value of h_0. The least squares' solution of these three points is

$$e_m = 4.7696 \pm 0.0041 + (0.00045 \pm 0.00686)m. \tag{55}$$

[39] See G.C. 1929, pp 36–40.

The ratio $r_e/r_i = 0.389$, indicating remarkable consistency of the data. The errors in Eq. (55) are accordingly based on r_i, and it is worthy of notice that if they were based on r_e they would be only 39 percent as large. Eq. (55) is plotted as curve h of Fig. 8, and the various resulting constants are listed in Table II. This figure shows also the curve necessary if $1/\alpha = 137$ and $e_{3/3}$ has the value given above. The intercept at $m = 0$ is 4.7824, as already given in connection with Eq. (45).

The values of e and h resulting from curve h are for all practical purposes identical with those adopted in 1929. In order to facilitate comparison, I give in Table II, on the next line below the curve h values, the 1929 constants. With the probable exception of curve k, to be discussed later, this curve h solution represents, I believe, the most probable values to be obtained from

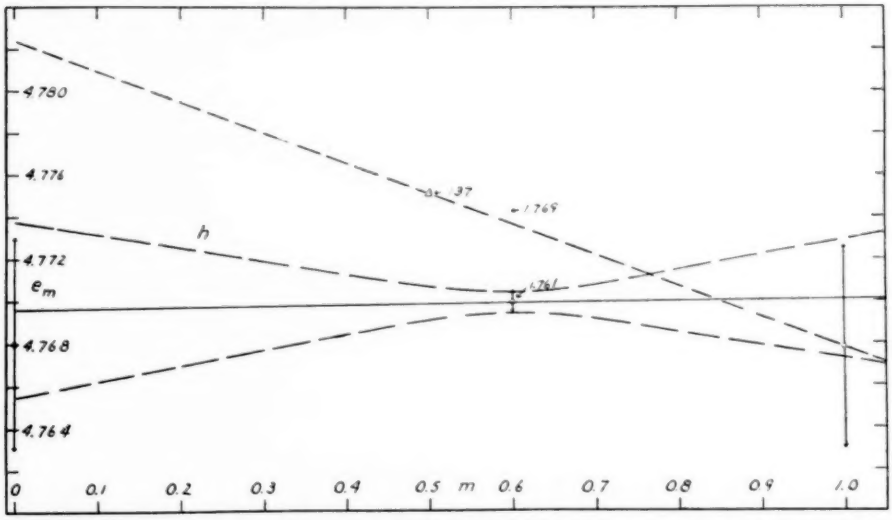

Fig. 8. Solution of Eq. (55). Data of Fig. 7, transformed to e_m values, plus the directly determined $e_d = 4.768$, at $m = 0$. $h = 6.5464$, $e = 4.7696$. (The letter h should have been placed on the continuous line).

the data at the present time. The chief difference between the 1929 and the 1932 values lies in the stated probable errors. As has already been noted, the 1929 errors were calculated in a very crude way, and as shown in more detail in the concluding section, some of them are too large, and some are too small.

(2) Curve *i*. It is now of interest to calculate the values of the constants that should have been obtained in 1929, had they been correctly calculated. For this purpose we use merely the data of curve e (Table III, col. 4), transformed to e_m values, plus $e_d = 4.770 \pm 0.005$ ($p = 0.40$), since that is the 1929 adopted value from all direct determinations of e. The $e_{3/5}$ point is the same as for curve h. The other points are

$$e_{3/4} = 4.7759 \pm 0.0036 \ (p = 0.77)$$

$$e_{3/3} = 4.7699 \pm 0.0034 \ (p = 0.87).$$

From these four points we obtain

$$e_m = 4.7688 \pm 0.0035 + (0.0021 \pm 0.0058)m. \quad (56)$$

This solution is plotted as curve i, Fig. 9.

The ratio $r_e/r_i = 0.76$ and the data are accordingly, as a whole, quite consistent, although the $e_{3/4}$ point is obviously too high, just as the $h_{4/3}$ point in curve e is low. The various constants calculated from Eq. (56) are listed in Table II, where they may be compared with the published 1929 values placed directly above them. It is of interest to note that the best present values, as given by curve h, agree rather better with the published 1929 values than with the values that should have been obtained in 1929. However, all three sets of values agree with each other far within the limits of error, and it is quite immaterial which set is used.

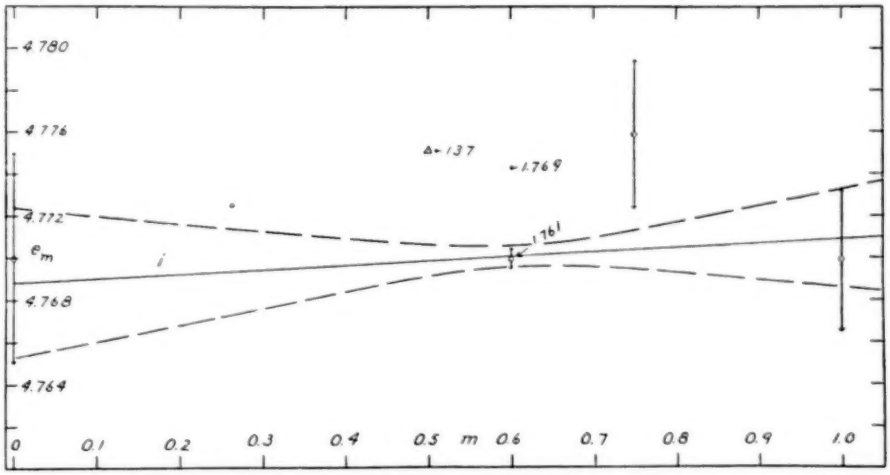

Fig. 9. Solution of Eq. (56). Data of Fig. 5, transformed to e_m, plus $e_d = 4.770$. The resulting constants ($h = 6.5441 \pm 0.0079$, $e = 4.7688 \pm 0.0035$) are the values that should have been obtained in 1929, had the calculations been made correctly.

New Solutions with X-ray Method at $n = 4/3$

The calculations outlined above were presented to the American Physical Society, at the Berkeley meeting, Dec. 1931.[40] Since then two more solutions have been made, based on the very recent work of Bearden.[41] It has been noted, at the end of the discussion of curve f, that that curve furnished independent evidence that the true x-ray wave-lengths are *not* those given by ruled gratings. Bearden has independently been led to this same conclusion from an entirely different source. Using the measured index of refraction of quartz, for various wave-lengths, he has been able to *calculate* values of the respective wave-lengths. In order to do this it is necessary to assume values of various general constants, including e/m. For this last constant Bearden

[40] R. T. Birge, Phys. Rev. **39**, 547 (1932) abstract 6.

[41] J. A. Bearden, Phys. Rev. **39**, 1 (1932), and further details by private communication.

adopted[42] 1.761. The resulting wave-lengths show remarkable agreement with those obtained by crystals, and are in definite disagreement with those obtained by ruled gratings. To get agreement with the latter, it would be necessary to adopt a value of e/m some 0.5 percent *lower* than 1.761.

I do not feel competent to make a critical estimate of the possible errors in the theoretical equation used by Bearden. If it is a sufficiently correct equation, and if the various constants appearing in it have been correctly evaluated, we are led to the surprising result that it is the grating wave-lengths that are wrong, and not the crystal wave-lengths, contrary to all previous opinion.

(1) Curve *j*. With this new condition of affairs, it is of interest to leave the x-ray value of h_n at $n=4/3$, and to adopt the crystal wave-lengths, just as was done in 1929. I have accordingly recalculated curve *f*, with this change. For $n=3/3$ we now have left three methods, with a weighted average $h_{3/3}=6.5500\pm0.0066$, as adopted for curve g. For $n=4/3$ there are two methods. The x-ray value of h_n is now 6.5526 ± 0.0063. This is the quantity to which 0.01740 was added, when transferring the x-ray observation from $n=4/3$ to $n=3/3$, in curve *f*. This x-ray observation is now to be averaged with $h_n=6.539\pm0.0042$[43] as obtained from σ. The weighted average is $h_{4/3}=6.5431\pm0.0042$ ($p=0.57$), and $r_e/r_i=1.21$. The expected deviation from unity is 34 percent, and so we have here no definite indication of inconsistency in the two methods. The $h_{5/3}$ point is as usual 6.5471 ± 0.0012 ($p=6.94$), based on $e/m=1.761\pm0.001$.

The solution of these three points is

$$h_n' = 6.5443 \pm 0.0133 + (0.00162 \pm 0.0081)n \qquad (57)$$

with $r_e/r_i=0.685$ and with the errors consequently based on r_i. This is certainly a far different situation from that found in curve *f*, where $r_e/r_i=2.14$, but the improvement is partly an illusion. It merely happens that the x-ray value of h_n, which even without the 0.017 correction is rather high, is here averaged with the low value of h_n derived from σ, to give a point $h_{4/3}$ that agrees well with the other two points. In curve *f* the low $h_{4/3}$ point came from σ alone, and the very high x-ray value helped to raise the resulting $h_{3/3}$ point and to produce the inconsistency of all points. It is certainly true, however, that the retention of the x-ray value of h at $n=4/3$ improves the consistency of the data, and there is now no sufficient reason for rejecting any one of the six methods.

Eq. (57) is plotted as curve *j*, Fig. 10, and the derived constants are listed in Table II. They are almost identical with the constants given by curve *i*. The value of e is 4.7688 ± 0.0059, and with the assumption that the crystal wave-lengths of x-rays are the correct values, this becomes the best inde-

[42] In previous work on the index of refraction of x-rays, by Stauss, Bearden and others, the wave-lengths as given by gratings were assumed to be correct, and a value of e/m was calculated.

[43] Here we use the correct probable error, of which mention has been made in footnotes 22 and 37.

pendent evaluation of e by the new method. It is really remarkable that it should agree so well with the oil-drop value, and this again is confirmation of Millikan's work.

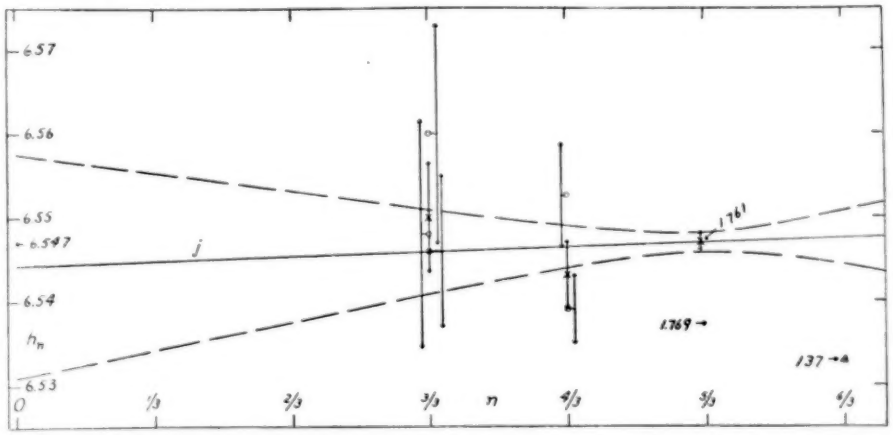

Fig. 10. Solution of Eq. (57). Data of Fig. 6, but with x-ray observation transferred back to $n=4/3$. $h=6.5443\pm0.0133$, $e=4.7688\pm0.0059$. This is the best value of e, as calculated by the new method.

(2) Curve *k*. The last solution to be presented is based on the data of curve j, converted to e_m values, and with the addition of $e_d=4.768\pm0.005$. The needed data are just those of curve h, with the added point $e_{3/4}=4.7738$

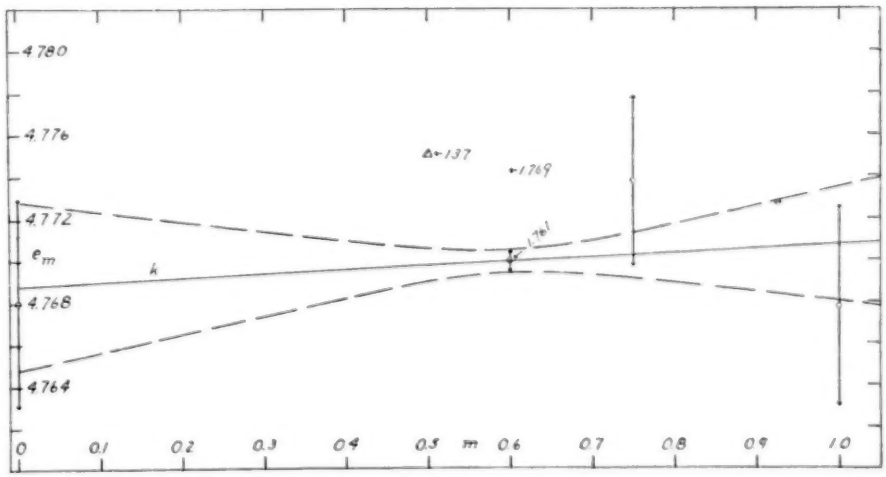

Fig. 11. Solution of Eq. (58). Data of Fig. 10, converted to e_m values, and $e_d=4.768$ added. The resulting constants ($h=6.5443\pm0.0091$, $e=4.7688\pm0.0040$) are considered the most reliable values, on the basis of present data.

±0.0041 ($p=0.60$), from $h_{4/3}=6.5431\pm0.0042$. The solution of the four points is

$$e_m = 4.7688 \pm 0.0040 + (0.00196 \pm 0.0066)m \qquad (58)$$

for which $r_e/r_i=0.51$. The equation is plotted as curve k, Fig. 11, and the

resulting constants are given in Table II. They are nearly identical with those obtained from curve *i*. These numerical coincidences are devoid of scientific significance, but they are nevertheless very interesting. Thus if one assumes that the x-ray method belongs at $n=3/3$, and must be ignored, together with the σ method, because of inconsistency, then the best resulting set of constants is given by curve *h*, and this set is almost identical with the constants published in 1929. On the other hand, if one assumes that the x-ray method belongs at $n=4/3$, the resulting set of constants is given by curve *k*, and this set is almost identical with the constants that would have been obtained in 1929, if correctly calculated (curve *i*).

General Discussion and Conclusions

Table II shows, in summary form, the various sets of constants resulting from the data and curves plotted in Figs. 1 to 11. The most interesting of these constants is probably $1/\alpha$. The source of interest lies in the fact that $1/\alpha$ is one of the very few dimensionless ratios formed from general physical constants. Its numerical value should therefore have theoretical significance, and the theories that have already been proposed in this connection are tabulated in *G. C.* 1929, pp. 71–72, and by Bond,[2] pp. 995–996. Of these theories the one most deserving of serious consideration seems to be that put forward by Eddington,[10] which predicts $1/\alpha=137$.

The tabular material of Table II shows that if $1/\alpha=137$, then e/m equals 1.768 or more. The broken line drawn in Figs. 7 and 8 and given by Eq. (45) corresponds to $e/m=1.7679$, as already noted. The only solution of the experimental material yielding a value of $1/\alpha$ equal to 137 within the probable error is that given by curve *a*, Fig. 1. This solution is based in Bond's 36 observations, yielding 3 points, with each point weighted according to the number of observations composing it. The point at $n=5/3$ comes from 13 observed values of e/m, of which only two equal 1.761. The other values are all higher, running up to 1.773. In fact the first observed "low" value of e/m was that by Babcock[44] in 1923. Since that date no new experimental value of e/m, with the exception of the magnetic deflection work by Wolf,[45] has been higher than 1.761. The most important conclusion of Bond's work is that $1/\alpha$ is really 137, from purely experimental evidence. We see now that this conclusion was reached only by using every value of e/m listed in certain compilations, and by giving the same weight to each investigation. With numerous older "high" values of e/m Bond thus got a high average. At the present time one may say with some assurance that either there is a serious constant error of essentially the *same* magnitude in the recent determinations of e/m by three totally different methods (Zeeman effect, fine structure of hydrogen and helium spectral lines, and acceleration of free electrons by an electric field), or else $1/\alpha$ is *not* 137. The two most probable sets of data (curves *h* and *k*) give 137.305 ± 0.048 as the average value. These solutions

[44] H. D. Babcock, Astrophys. J. **58**, 149 (1923). $e/m=1.761$ from Zeeman effect. See page 44 of G.C. 1929.

[45] F. Wolf, Ann. d. Physik **83**, 849 (1927). See p. 43, *G.C.* 1929.

have been obtained on the assumption that the experimental value of e/m equals 1.761 ± 0.001, and the value calculated from the solutions is identical with this, as shown in Table II.

With such an assumption regarding e/m, the resulting values of e and h fall within certain rather narrow limits. These limits may be brought out more clearly by a brief review of the various solutions that have been obtained. Curve b results from Bond's 36 observations, but with the three points weighted according to their probable errors. Curve c is based on the same method of weighting but with the addition of three new determinations of e/m. The resulting values of e and h are slightly higher than those advocated in 1929, but this result is due to the relatively high averate value of e/m obtained from the inclusion of all the older work. Solution d is based on an arbitrary equal weighting of the six observations adopted in 1929, and has no real merit. A more logical weighting of these same six observations gives the quite different solution shown by curve e, and this solution represents the values that should have been obtained in 1929, as based on Eq. (1). Since, however, Eq. (1) does not include any directly determined value of e, the solution merely shows that the 1929 data give a value of e (4.765) by the new method that is quite consistent with the oil-drop value (4.768).

A re-examination of all available data then leads to curve f, which however could not be used because of definite inconsistency of the data ($r_e/r_i = 2.14$). Solution g, obtained by omitting the x-ray and σ methods, is properly self-consistent and the resulting value $e = 4.7732 \pm 0.0072$ represents the best independent determination of e by the new method, *provided* we assume that the x-ray observation belongs at $n = 3/3$. The use of Eq. (2) and the addition of the direct determination $e = 4.768$, then gives curve h, which is one of our two most probable solutions. The other of these most probable solutions is curve k, and this differs so little from curve h that it is really immaterial which is accepted. Curve k results from the assumption that the crystal wave-lengths of x-rays are correct, an assumption that places the x-ray method at $n = 4/3$. Two independent sources of evidence,—Bearden's recent work,[41] and the increased consistency of the data resulting from the transfer of the x-ray method from $n = 3/3$ to $n = 4/3$,—unite to indicate the correctness of the assumption. This was discussed at the opening of the section entitled "New Solutions with X-ray Method at $n = 4/3$." Curve j is the solution based on Eq. (1), corresponding to curve k, based on Eq. (2). Curve i is the solution, based on Eq. (2), of my 1929 data, and represents the values that should have been obtained at that time.

This brings us to the final conclusions resulting from our analysis of the data. The crystal x-ray wave-lengths are probably correct, and curve k represents the most reliable solution. The best value of e, obtained by the new method, is given by curve j, viz. $e = 4.7688 \pm 0.0059$. The best final value of e is 4.7688 ± 0.0040, from curve k. As was stated in the published abstract of this paper,[40] the possibility of the simultaneous evaluation of e and h, from the totality of experiments designed to measure h, is of the utmost importance. The fact that the new value of e thus obtained is almost identical with

the oil-drop value is, in my opinion, very strong evidence that there is no unsuspected error of any significance in Millikan's work. On the other hand, the preceding analysis has shown that the most probable values of e, h and $1/\alpha$ depend primarily on the value adopted for e/m. The present evidence favors $e/m = 1.761$, and with this one gets, almost inevitably, a value of e agreeing well with the oil-drop value, and a value of $1/\alpha$ definitely different from 137. If $1/\alpha$ is to equal 137, it is necessary that e/m equal approximately 1.768. As discussed in connection with Eq. (45), the resulting values of e and h are then about 0.26 percent and 0.30 percent higher than the values adopted in 1929. The direct and necessary dependence of the values of e, h and $1/\alpha$ on the value adopted for e/m is the most important result brought out in this paper.

With curve k assumed as the most probable solution, the resulting values of e, h, $1/\alpha$ and e/m differ so little from those suggested by the writer in 1929 that, for practical purposes, *no* change is advocated. It would seem good policy to recommend a change in the value of a fundamental physical constant *only* when the new value differs from the old by more than the probable error of the new value. When two values agree within less than the probable error, their difference has no scientific significance. In view of the wide use of fundamental constants, and of the numerous constants that may be derived from them[46] it is desirable, as a purely practical matter, that the values in common use be changed as infrequently as possible. This has been the policy of all atomic weight committees. The writer feels that, as a result as much of good luck as good management, he advocated in 1929 values of e, h, e/m and $1/\alpha$ that are still quite acceptable. In special tests of theoretical relations one should of course use the latest and most reliable values of all constants. For such work the constants listed on the last line of Table II (solution k) are recommended, but for all ordinary work the 1929 values may well be retained.

The only significant error in these 1929 values lies in the stated probable error of all ratios of the type h^x/e^y. The method adopted in 1929 for calculating the error in such cases is outlined in footnote 2, page 60 of G. C. 1929. The preceding discussions of the present paper have shown the correct method. Thus, from Eqs. (14) and (15)

$$\frac{2\pi}{\alpha c} = \frac{h}{e^2} = \frac{h_2}{{e_0}^2} . \tag{59}$$

The proportional error in h/e^2, or in α, is merely the proportional error in h_2, as calculated from the adopted $h_n : n$ curve. Similarly, from Eq. (53)

$$\frac{2\pi}{\alpha c} = \frac{h_0}{{e_{1/2}}^2} \tag{60}$$

and the proportional error in α is twice the proportional error in $e_{1/2}$, as calculated from the adopted $e_m : m$ curve. This last curve always yields the most

[46] Table c of G.C. 1929 gives an incomplete list of such derived constants.

reliable values of e and h. Hence Eq. (60) is to be used, rather than (59). We may combine and generalize Eqs. (59) and (60) as follows,

$$\frac{h}{e^n} = \frac{h_n}{{e_0}^n} = \frac{h_0}{{e_{1/n}}^n} \cdot \tag{61}$$

Replacing n by its equivalent $1/m$ and forming the mth power, Eq. (61) becomes

$$\frac{h^m}{e} = \frac{{h_{1/m}}^m}{e_0} = \frac{{h_0}^m}{e_m} \cdot \tag{62}$$

This last equation is most convenient for use with the $e_m{:}m$ curve, and states that the ratio h^m/e has a proportional error equal to the proportional error in the calculated value e_m. Similarly Eq. (61) states that the proportional error in h/e^n equals that in the calculated h_n. Writing

$$h^x/e^y = (h^m/e)^y = (h/e^n)^x \tag{63}$$

where $m = x/y$, or $n = y/x$, we obtain from Eqs. (61) (62) and (63) the final generalization that the proportional probable error in h^x/e^y is y times the proportional probable error in e_m, as obtained from the adopted $e_m{:}m$ curve, or x times the proportional probable error in h_n, as obtained from the adopted $h_n{:}n$ curve. The $e_m{:}m$ curve includes the most data and is therefore usually the more reliable.

As examples of the use of Eq. (63) we may note that the correct probable error in h/e^2 (or in $1/\alpha$) is less than one-half as large[47] as that given in 1929. On the other hand, the ratio h/e is, from solution k, $(1.3723 \pm 0.0008) \times 10^{-17}$ in place of 1.3725 ± 0.0005 given in 1929. The error is thus larger than before. Many of the constants listed in Tables b and c of G.C. 1929 contain the factor h^x/e^y, and the probable errors in all such cases require more or less revision.[48] It is, however, not my purpose here to present any extended revision of the general physical constants.[49]

This paper may well conclude with the remark made in the introduction,—that the primary object at this time is to present a mathematically correct method for the (necessarily) simultaneous evaluation of e and h and of the probable errors in all ratios of the type h^x/e^y. It is only in respect to these points that, so far as I am now aware, the methods used in G.C. 1929 are open to real criticism.

[47] 137.307 ± 0.048, in place of $137.29_4 \pm 0.11$.

[48] The method used in 1929 involved the assumption of a certain *constant* error for h_n or e_m, instead of the varying error shown in Figs. 1 to 11 of this article. Hence the true probable error, for values of m or n near the "center of gravity" of the data, is *less* than that given in 1929, while for more distant values it is greater.

[49] There is now in existence a National Research Council Committee on Physical Constants, of which the writer is chairman. This committee will issue a detailed report whenever any significant changes in the *values* (not the probable errors) of the fundamental constants seem required.

A RELIABLE METHOD OF OBTAINING THE DERIVATIVE FUNCTION FROM SMOOTHED DATA OF OBSERVATION*

By George Rutledge
Massachusetts Institute of Technology

(Received February 29, 1932)

Abstract

In this process of differentiation there is involved no attempt to determine empirically an equation capable of representing the given data sufficiently well to admit differentiation. On the contrary, a differentiating tool is applied successively to various small sections of the data. This tool is the fourth degree polynomial determined by five points.

1. Introduction

THE method here described for finding, *with known limit of error*, the derivative function of a differentiable function, from values of the latter observed for equally-spaced values of the independent variable, was originally developed for the specific purpose of finding the derivative function dp/dt of the function defined by the vapor pressure values of steam.[1] These vapor pressure values had been observed at intervals of five degrees and smoothed by the usual method of examining successive differences. In the case of steam it was possible to carry the smoothing to fourth differences (not constant, but monotone). The primary data in this instance were undoubtedly of somewhat unusual precision (one part in five thousand). For values of t at intervals of ten degrees from zero degrees to three hundred seventy degrees, the derivative function dp/dt, obtained from the smoothed values of p by the process to be described in this paper, is the following:

Table I. *dp/dt at saturation for water, in atmospheres per degree centigrade.*

t	dp/dt	t	dp/dt
100	0.03571	240	0.5803
110	0.04752	250	0.6623
120	0.06210	260	0.7521
130	0.07979	270	0.8501
140	0.10096	280	0.9565
150	0.12596	290	1.0719
160	0.15517	300	1.1969
170	0.18892	310	1.332
180	0.22754	320	1.480
190	0.27136	330	1.640
200	0.3207	340	1.815
210	0.3760	350	2.005
220	0.4375	360	2.214
230	0.5055	370	2.444

* Presented to the American Mathematical Society, March 25, 1932.

[1] F. G. Keyes and L. B. Smith, "Some Final Values for the Properties of Saturated and Superheated Water," *Mechanical Engineering* **53**, 132 (1931).

The method of computation to be described automatically determines the number of figures which are of value, since it consists in retaining the figures common to three determinations, the errors in which are normally not all of the same sign. In the final section of this paper rigorous limitations upon the size of the error in these determinations are established.

The values of p may be recovered from the above table by use of Simpson's rule or by use any one of the more accurate formulas for approximate integration (for example, the values of p for $t=100, 140, 180, 260, \ldots$ which are, respectively, 1.000, 3.567, 9.895, 22.881, 46.286, . . .), thus directly justifying the function dp/dt as recorded.

The writer has been assured by Professor Keyes that many scientific workers would welcome a simple and reliable method of obtaining a derivative function from data comparable with the data of Keyes and Smith. Even where it has been possible to represent such data by an empirical formula, differentiation of this formula has been unsatisfactory as a means of determining the derivative function.

Formulas for the derivative of the interpolation polynomial in one or another of the classic forms of Newton, Stirling, and Gauss, may be found in any good book[2] on interpolation and numerical differentiation and integration.

It is only recently, however, that the interpolation polynomial has been studied in a form free from differences.[3] Tables for interpolation without the use of differences, based on this form of the polynomial, have been published[4] recently, and the advantages, both theoretical and practical, of tables of this form have been pointed out.

A natural extension of this development is the presentation of a table for differentiating without the use of differences. This is the purpose of the present paper.

2. The Interpolation Polynomial and its Derivative

The table to be presented will be based on a fourth degree polynomial determined by five points, but this polynomial is merely a tool to be carried from place to place, so to speak, and there is no implication that the data as a whole can be adequately represented by a polynomial or any other elementary type of function.

It is easy to verify that the polynomial $P(\lambda)$, of degree four[5] (at most) in λ, which for $\lambda=-2, -1, 0, 1, 2$ takes, respectively, the values y_{-2}, y_{-1}, y_0, y_1, y_2 is the following:

[2] For example, Runge-König, Vorlesungen über numerisches Rechnen, or Whittaker and Robinson, Calculus of Observations.

[3] Rutledge, *Transactions of the American Mathematical Society*, **26**, 113–123 (1924); **31**, 807–820 (1930); also *Journal of Mathematics and Physics of the Massachusetts Institute of Technology* **2**, 47–62 (1922); **8**, 1–12, 13–17 (1929); **9**, 166–180 (1930).

[4] Huntington, *Proceedings of the American Academy of Arts and Sciences*, **63**, 421 (1929); also Rutledge (and Crout), reference 3., *Journal*, **9** (1930).

[5] For the general form of $P(\lambda)$ of degree $2n$, see Rutledge, reference 3., *Journal*, **2** (1922), and *Transactions*, **26** (1924).

$$
\begin{aligned}
P(\lambda) = {} & y_0 \\
& + \frac{1}{4!}\,(2y_{-2} - 16y_{-1} \qquad + 16y_1 - 2y_2)\lambda \\
& + \frac{1}{4!}(-\,y_{-2} + 16y_{-1} - 30y_0 + 16y_1 - y_2)\lambda^2 \\
& + \frac{1}{4!}(-\,2y_{-2} + \;4y_{-1} \qquad - \;4y_1 + 2y_2)\lambda^3 \\
& + \frac{1}{4!}\,(y_{-2} - \;4y_{-1} + \;6y_0 - \;4y_1 + y_2)\lambda^4.
\end{aligned} \tag{1}
$$

The derivative $P'(\lambda)$ has the form

$$\frac{1}{12}(C_{-2}y_{-2} + C_{-1}y_{-1} + C_0y_0 + C_1y_1 + C_2y_2), \tag{2}$$

where, for $\lambda = -1, 0, 1$, the C's are the integers given in Table II.

TABLE II.

λ	C_{-2}	C_{-1}	C_0	C_1	C_2
−1	−3	−10	+18	− 6	+1
0	+1	− 8	0	+ 8	−1
+1	−1	+ 6	−18	+10	+3

In order to illustrate adequately the limits of error in the approximation process by which we determine values of a derivative function, we resort to data involving a much larger number of significant figures than physical or chemical observation will ordinarily yield. The principles involved are unaffected by this circumstance. The data of the following table may still be regarded as inexact data of observation. In fact the ten-place values given are obtained by rounding off a twenty-three-place table[6] to ten places of decimals.

TABLE III.

$P_\alpha(\lambda)$ λ	$P_\beta(\lambda)$ λ	$P_\gamma(\lambda)$ λ	x	sin x
−2			0.7	0.64421 76872
−1	−2		0.8	0.71735 60909
0	−1	−2	0.9	0.78332 69096
+1	**0**	**−1**	**1.0**	**0.84147 09848**
+2	+1	0	1.1	0.89120 73601
	+2	+1	1.2	0.93203 90860
		+2	1.3	0.96355 81854

We shall illustrate the method of determining the derivative function by finding just one value of the derivative, namely, the derivative for $x = 1$.

[6] Van Orstrand, *Memoirs of the National Academy of Sciences*, **14** (1921), Fifth Memoir.

The function sin x is approximated in succession by each of three *different* polynomials in λ, namely, $P_\alpha(\lambda)$, $P_\beta(\lambda)$, $P_\gamma(\lambda)$, with correspondingly three different relations between x and the independent variable of the polynomial approximation, as shown in Table III.

By use of the coefficients of Table II we obtain

$$\begin{aligned} P_\alpha'(+1) &= 0.05403\ \ 053\cdots, \\ P_\beta'(0) &= 0.05403\ \ 005\cdots, \\ P_\gamma'(-1) &= 0.05403\ \ 046\cdots, \end{aligned} \tag{3}$$

with *residual errors*

$$\begin{aligned} \epsilon_\alpha &= +0.00000\ \ 030\cdots, \\ \epsilon_\beta &= -0.00000\ \ 018\cdots, \\ \epsilon_\gamma &= +0.00000\ \ 203\cdots, \end{aligned} \tag{4}$$

since the true value of the derivative approximated is known to be 0.05403 023 . . . . These residual errors would be unaffected by further refinement of the data, for example, by using twenty-three places of decimals instead of ten. Equally, the residual errors, by definition, are unaffected by coarsening the data, for example, by using four or five places of decimals instead of ten.

It is worthy of comment at this point that the results of Eq. (3) are each obtainable by a continuous series of operations with a computing machine with no necessity for the recording of intermediate results.

If we shorten Table III to five places (for convenience we use the first five figures of each value of sin x, without rounding off) we obtain a table comparable with a table of physical data. The *errors of observation* are then $-0.00000\,8$, $-0.00000\,6$, $-0.00000\,7$, $-0.00000\,1$, $-0.00000\,7$, $-0.00000\,9$, $-0.00000\,8$. The changes which these errors of observation will make in the computed values of $P_\alpha'(+1)$, $P_\beta'(0)$, $P_\gamma'(-1)$, are dependent on these errors alone and on the coefficients of Eq. (2). These changes are of the form (the δ's indicating the above mentioned errors)

$$\frac{1}{12}\sum_{i=-2}^{2} C_i\delta_i \quad . \tag{5}$$

and hence are of the same order of magnitude as the errors of observation. With Table III shortened as stated above, we obtain therefore

$$\begin{aligned} P_\alpha'(+1) &= 0.054036, \\ P_\beta'(0) &= 0.054030, \\ P_\gamma'(-1) &= 0.054026, \end{aligned} \tag{6}$$

from which we conclude that the value 0.05403 is correct to five places of decimals. The derivative of sin x with respect to x, for $x=1$, is then 0.5403.

This is the method by which the number of figures to be retained for the values of dp/dt tabulated in Table I was in each case determined. Each result tabulated consists of the figures common to $P_\alpha'(+1)$, $P_\beta'(0)$, $P_\gamma'(-1)$.

Thus the effect of errors of observation is eliminated. Agreement between $P_\alpha'(+1)$, $P_\beta'(0)$, $P_\gamma'(-1)$ will be poor if the errors of observation are large or extremely erratic. The question of appropriateness of this process for a given set of data is therefore automatically settled.

3. The Residual Errors

Let $f(x)$ and six derivatives, $f'(x), f''(x), \cdots, f^{vi}(x)$ be defined on the interval $-2 \leqq x \leqq 2$. Consider five constants

$$-2 \leqq c_1 < c_2 < c_3 < c_4 < c_5 \leqq +2. \tag{7}$$

Let $x = \lambda\theta$, $0 \leqq \theta \leqq 1$, and approximate $f(x)$ in the interval $-2\theta \leqq \lambda\theta \leqq +2\theta$ by a polynomial of degree four (at most) in λ, so determined that it takes the values

$$f(c_1\theta),\ f(c_2\theta),\ f(c_3\theta),\ f(c_4\theta),\ f(c_5\theta) \tag{8}$$

for $\lambda = c_1, \cdots, c_5$, respectively. The derivative of this polynomial with respect to λ, for any given value of λ, has the form

$$K_1(\lambda)f(c_1\theta) + K_2(\lambda)f(c_2\theta) + \cdots + K_5(\lambda)f(c_5\theta), \tag{9}$$

where the K's are coefficients, illustrated by Table II of Part 2, dependent on λ, but not dependent on $f(x)$. The derivative of $f(x)$ with respect to λ, for the same given value of λ, is $\theta \cdot f'(\lambda\theta)$. Thus far we have regarded the parameter θ as fixed. Now let θ vary and consider $\phi(\theta)$, $\phi'(\theta)$, $\cdots$, defined as follows:

$$\begin{aligned}
\phi(\theta) &= \theta f'(\lambda\theta) - \sum K_i(\lambda) f(c_i\theta), \\
\phi'(\theta) &= f'(\lambda\theta) + \lambda\theta f''(\lambda\theta) - \sum K_i(\lambda) c_i f'(c_i\theta), \\
\phi''(\theta) &= 2\lambda f''(\lambda\theta) + \lambda^2\theta f'''(\lambda\theta) - \sum K_i(\lambda) c_i^2 f''(c_i\theta), \\
\phi'''(\theta) &= 3\lambda^2 f'''(\lambda\theta) + \lambda^3\theta f^{iv}(\lambda\theta) - \sum K_i(\lambda) c_i^3 f'''(c_i\theta), \\
\phi^{iv}(\theta) &= 4\lambda^3 f^{iv}(\lambda\theta) + \lambda^4\theta f^{v}(\lambda\theta) - \sum K_i(\lambda) c_i^4 f^{iv}(c_i\theta), \\
\phi^{v}(\theta) &= 5\lambda^4 f^{v}(\lambda\theta) + \lambda^5\theta f^{vi}(\lambda\theta) - \sum K_i(\lambda) c_i^5 f^{v}(c_i\theta).
\end{aligned} \tag{10}$$

If we let θ approach zero, we obtain:

$$\begin{aligned}
\phi(0) &= \{\ \ - \sum K_i(\lambda)\} \cdot f(0), \\
\phi'(0) &= \{1 - \sum c_i K_i(\lambda)\} \cdot f'(0), \\
\phi''(0) &= \{2\lambda - \sum c_i^2 K_i(\lambda)\} \cdot f''(0), \\
\phi'''(0) &= \{3\lambda^2 - \sum c_i^3 K_i(\lambda)\} \cdot f'''(0), \\
\phi^{iv}(0) &= \{4\lambda^3 - \sum c_i^4 K_i(\lambda)\} \cdot f^{iv}(0).
\end{aligned} \tag{11}$$

We also obtain

$$\phi^{v}(0) = \{5\lambda^4 - \sum c_i^5 K_i(\lambda)\} f^{v}(0), \tag{12}$$

but we are not concerned with the latter.

Now if $f(x)$ is a polynomial of degree four, $\phi(\theta)$ is identically zero, and since $f(0)$, $f'(0)$, $f''(0)$, $f'''(0)$, $f^{iv}(0)$ may all be different from zero, it results that the K's must satisfy the following system of linear equations:

$$\begin{aligned} \sum K_i(\lambda) &= 0, \\ \sum c_i K_i(\lambda) &= 1, \\ \sum c_i^2 K_i(\lambda) &= 2\lambda, \\ \sum c_i^3 K_i(\lambda) &= 3\lambda^2, \\ \sum c_i^4 K_i(\lambda) &= 4\lambda^3, \end{aligned} \tag{13}$$

From this it follows that, *whatever the given function $f(x)$ may be*, provided as stated initially, that certain derivatives of $f(x)$ exist,

$$\phi(0) = \phi'(0) = \phi''(0) = \phi'''(0) = \phi^{iv}(0) = 0. \tag{14}$$

In general, $\phi^v(0) \neq 0$, but we have, from Eq. (10),

$$|\phi^v(\theta)| \leqq \{5|\lambda|^4 + \sum |K_i(\lambda)| \cdot |c_i|^5\} f_{max.}^{\ v} + \theta \cdot |\lambda|^5 f_{max.}^{\ vi}, \tag{15}$$

which may be written briefly

$$|\phi^v(\theta)| \leqq \sigma_1 f_{max.}^{\ v} + \theta \cdot \sigma_2 f_{max.}^{\ vi}, \tag{16}$$

where $f_{max.}^{\ v}$ and $f_{max.}^{\ vi}$ denote the maximum absolute values of $f^v(\lambda\theta)$ and $f^{vi}(\lambda\theta)$, for $-2 \leqq \lambda \leqq 2$. Now we have, since $\phi^{iv}(0) = 0$,

$$\begin{aligned} |\phi^{iv}(\theta)| &= |\phi^{iv}(\theta) - \phi^{iv}(0)| = \left| \int_0^\theta \phi^v(\theta) d\theta \right| \\ &\leqq \int_0^\theta |\phi^v(\theta)| \, d\theta \leqq \theta \cdot \sigma_1 f_{max.}^{\ v} + \frac{\theta^2}{2} \cdot \sigma_2 f_{max.}^{\ vi}. \end{aligned} \tag{17}$$

Repeating this process we obtain

$$\begin{aligned} |\phi'''(\theta)| &\leqq \frac{\theta^2}{2!} \cdot \sigma_1 f_{max.}^{\ v} + \frac{\theta^3}{3!} \cdot \sigma_2 f_{max.}^{\ vi}, \\ &\cdots\cdots\cdots\cdots\cdots, \\ |\phi(\theta)| &\leqq \frac{\theta^5}{5!} \cdot \sigma_1 f_{max.}^{\ v} + \frac{\theta^6}{6!} \cdot \sigma_2 f_{max.}^{\ vi}. \end{aligned} \tag{18}$$

Finally we place $\theta = 1$, $\lambda = \pm 1$, $c_1 = -2$, $c_2 = -1$, $c_3 = 0$, $c_4 = 1$, $c_5 = 2$, and, from Table III, $\Sigma |K_i(\lambda)| \cdot |c_i|^5 = 12$.

Thus we obtain

$$|\phi(1)| \leqq 17 \frac{f_{max.}^{\ v}}{5!} + \frac{f_{max.}^{\ vi}}{6!}, \tag{19}$$

where, since $\lambda = x$ when $\theta = 1$, the derivatives are with respect to λ.

Similarly, with $\lambda = 0$, we obtain

$$|\phi(1)| \leqq \frac{20}{3} \frac{f_{\max.}^{v}}{5!}. \tag{20}$$

Applying the foregoing limits to the residual errors Eq. (4) we have

$$\begin{aligned} |\epsilon_\alpha| &\leqq \frac{17}{120}(0.00000\ \ 76) + \frac{1}{720}(0.00000\ \ 06), \\ |\epsilon_\beta| &\leqq \frac{1}{18}(0.00000\ \ 70), \\ |\epsilon_\gamma| &\leqq \frac{17}{120}(0.00000\ \ 62) + \frac{1}{720}(0.00000\ \ 08), \end{aligned} \tag{21}$$

that is,

$$\begin{aligned} |+0.00000\ \ 030| &< 0.00000\ \ 10, \\ |-0.00000\ \ 018| &< 0.00000\ \ 04, \\ |+0.00000\ \ 023| &< 0.00000\ \ 09. \end{aligned} \tag{22}$$

In practice, differences may replace derivatives in the limits Eqs. (19) and (20). We are normally able on this basis to assure ourselves that the residual errors are smaller than the errors of observation. When this is the case, the accuracy of the differentiation method herein described is limited only by the accuracy with which the function to be differentiated can be measured and smoothed.

THE CALCULATION OF NORMAL VAPOR PRESSURES FROM THE DATA OF THE GAS CURRENT METHOD, PARTICULARLY IN THE CASE OF IODINE

By Harold T. Gerry and Louis J. Gillespie
Massachusetts Institute of Technology*

(Received February 20, 1932)

Abstract

The gas current method of determining the vapor pressure of slightly volatile substances is capable of a high degree of experimental precision and has often been applied. The question—whether the vapor pressure, so determined and subsequently corrected only for the Poynting effect, is accurately equal to the normal vapor pressure of the substance—has not hitherto been the subject of theoretical investigation. A method, dependent on recent developments in the thermodynamic treatment of mixtures of real gases is here developed for calculating the normal vapor pressure from the data of the gas current method. The suggested method requires data at varying total pressure for one inert gas at several temperatures and relatively accurate data at one atmosphere total pressure for two inert gases over the range of temperature. The one-atmosphere data for at least one of the inert gases must be of the precision desired on the normal vapor pressures. The method is applied to the case of the vapor pressure of solid iodine from 0 to 100°C, for which substance data exist which are suitable, though not obtained for the most appropriate choice of experimental conditions. The new corrections (for deviations from the laws of ideal gases) are found to be important numerically in the case of the best measurements. The errors inherent in the various methods of applying the gas current method are briefly discussed. A graphic, empirical, method is found liable to a characteristic source of error.

Introduction

A METHOD frequently employed for the determination of vapor pressures is the so-called gas current method, in which a stream of air or other gas is passed over the substance at constant temperature, causing it to volatilize, and the content of substance in the gas stream is determined. Precautions must be taken that the gas stream is just saturated with the substance at the known temperature. In the calculation of the vapor pressure small corrections are theoretically necessary, as is well understood, for any lowering of vapor pressure occasioned by solubility of the gas in the substance, and for the slight increase of vapor pressure due to the pressure of the gas. It does not appear to be universally understood that the calculations, as hitherto always carried out, assume laws of pure gases or of gaseous mixtures which are not exact.

The vapor pressure of a substance is defined as the pressure of the vapor (under a negligible influence of a gravitational or other external field) when it is pure and in equilibrium with the pure substance at a definite tempera-

* Contribution from the Research Laboratory of Physical Chemistry, Massachusetts Institute of Technology, No. 283.

ture. If the substance is under a pressure greater than that of the vapor, the equilibration takes place through a semipermeable membrane, and the pressure as well as the temperature must be specified. It is the object of the gas current method to determine the *normal* vapor pressure, i.e., the vapor pressure when the pressure is the same on liquid and vapor and both phases are pure.

The methods previously used for the calculation of vapor pressures from the data are in general of three types. One form in which the gas current method has very often been applied is based on the assumption that the vapor pressure under the experimental conditions of pressure and gas solubility is equal to the product of the total pressure of the gas stream times the mol fraction of the substance in the gas stream. This assumption includes the laws of Boyle and Avogadro.[1,2]

Another way of applying the gas current method is to determine the concentration of the substance in the gas stream, i.e., the mass of substance per unit of volume of the gas mixture. Ordinarily the assumption would then be made that this concentration is equal to the concentration of pure vapor in equilibrium with liquid when this contains the same quantity of dissolved gas as in the experiment and is under the same total pressure as then. If the volume of the gas mixture has been directly measured, which in general is a difficult procedure, this method of calculation does not assume the ideal gas laws in any way,[3] but assumes[2] Gibbs' form of Dalton's law, which under the usual conditions for gas current vapor pressures is likely to result in an error greater than that of the previous method, as will be shown later.

A graphical method of extrapolation of the results to zero pressure of the inert gas is uncertain, due to the fact that the deviations from the ideal gas law are functions not only of the pressure but also of the composition of the gas mixture. In the variation of the total pressure under the experimental conditions the mol fraction also varies, and in such a manner as to introduce probably a curvature at pressures below the usual experimental range. The nature of this difficulty in extrapolation will be further discussed in a later section.

From recent work on the thermodynamics of gas mixtures from the point of view of equations of state[4] we are able to calculate with a considerable degree of accuracy normal vapor pressures from the data of the gas current method, provided that certain additional data are available. This method does not require the assumption of certain inexact laws such as the ideal gas law, Dalton's law, or the fugacity rule of Lewis and Randall.

In the present instance we shall follow the low pressure treatment of

[1] Lurie and Gillespie, J. Am. Chem. Soc. **49**, 1146 (1927).

[2] Gillespie, Phys. Rev. **36**, 121 (1930).

[3] Except that the ideal gas law may be used, usually with no appreciable error, for calculating the normal vapor pressure from that under the pressure of the gas stream.

[4] See Gillespie and Beattie, Phys. Rev. **36**, 743 (1930), for a summary of previous work and for new evidence.

Gillespie[5] of the equilibrium pressure of a gas in a mixture, making certain slight modifications in the method of combination of constants which appear from recent unpublished work to give a better general agreement with the experimental data. By the use of a sufficient amount of data from the gas current method, approximate values of the three most important constants of the vapor under consideration in the Beattie-Bridgeman equation of state[6,7] can be obtained, and these are then used to calculate normal vapor pressures from the pressure, temperature and mol fraction data of the gas current method.

This method will then be applied to the data of Braune and Strassman[8] on iodine and extended to calculate corrections to other data on the vapor pressure of iodine.

The Calculation of Normal Vapor Pressures

The low pressure equation of Gillespie for a binary mixture is modified by the use of linear square root combination of c similar to that used for A_0 instead of linear combination of c as was previously used. This change seems justified in the light of recent unpublished work by Dr. Eli Lurie and by others in this laboratory. The effect of this change is small except in the case of gases whose c constants are large, where a markedly superior agreement with experiment is obtained. The resulting equation is

$$2.303 \log \frac{p_1}{px_1} = \left(B_{01} - \frac{A_{01}}{RT} - \frac{c_1}{T^3}\right)\frac{p - p_1}{RT} + \left[\frac{(A_{01}^{1/2} - A_{02}^{1/2})^2}{RT} + \frac{(c_1^{1/2} - c_2^{1/2})^2}{T^3}\right]\frac{px_2^2}{RT} \cdot \quad (1)$$

The constants B_{01}, A_{01}, and c_1 are the constants in the Beattie-Bridgeman equation for the vapor of the liquid or solid, A_{02} and c_2 are the corresponding constants in the same equation for the inert gas used, p is the total pressure of the gas mixture, p_1 the vapor pressure of the liquid or solid under the total pressure p and saturated with inert gas. R is the gas constant, 0.08206 liter atmospheres per degree per mol, T the absolute temperature ($t°C+273.13$) and x_1 the mol fraction of the vapor and x_2 that of the inert gas in the gas mixture. Common logarithms are denoted by log.

In general, p_1 will be sufficiently low in ordinary gas current vapor pressure measurements that the perfect gas volume may be used in calculating the normal vapor pressure from that under the experimental conditions, and the liquid or solid may also be considered incompressible in this correction. By a proper choice of inert gases the effect of solubility may usually be made small enough for Raoult's law of vapor pressure lowering to be sufficiently exact. The normal vapor pressure is then given by

[5] Gillespie, Phys. Rev. **34**, 1605 (1929) Eq. (37). In this paper a subscript 1 was unfortunately applied to the last p of the equation. This subscript should be omitted.

[6] Beattie and Bridgeman, Proc. Amer. Acad. Arts and Sci. **63**, 229 (1928).

[7] Beattie and Bridgeman, J. Am. Chem. Soc. **50**, 3133 (1928); Zeits. f. Physik **62**, 95 (1930).

[8] Braune and Strassman, Zeits. f. physik. Chem. **143A**, 225 (1929).

$$\log p_0 = \log p_1 - \frac{V_1(p - p_0)}{2.303RT} - \log y. \qquad (2)$$

V_1 is the molal volume of the solid or liquid in the condensed phase and y is the mol fraction of the liquid or solid in the liquid or solid phase.

Since the constants A_{02} and c_2 are known for practically all gases that are likely to be used as inert gases[6,7] and the volume V_1 and the mol fraction y are either known or readily determined to the accuracy required, the constants A_{01}, B_{01} and c_1 for the vapor are the only further quantities required for a calculation of the normal vapor pressure when p, x_1 and T are determined by experiment. If they are known the calculation proceeds readily in spite of the double occurrence of the unknown in both equations. The solution will of necessity proceed by successive trial and approximation but the convergence is very rapid in practice.

In general the necessary constants A_{01}, B_{01}, and c_1 will not be known and some method of evaluating them is necessary. This can be done by the use of suitable data and Eqs. (1) and (2). If the available data were of extraordinary accuracy and we could count implicitly on the form of the temperature functions in Eq. (1), then gas current vapor pressure data over a range of temperature and of varying pressure at each temperature using a single inert gas would be sufficient to solve for the desired constants. In practice, however, additional data must be used in determining them. Since the coefficient of $(p-p_1)$ is independent of the inert gas used, it is possible to obtain information on the coefficient of px_2^2 from additional data with a different inert gas.

The simplest method of carrying out such a computation requires the following data:

(1) Accurate values of p and the corresponding mol fractions over the desired temperature range with the given vapor in the gas stream for two different inert gases whose equation of state constants are known. These data need only be for one value of the pressure in each gas provided the pressures used in the two gases are the same at any given temperature. Since these data are those to be used for determining the normal vapor pressures, their accuracy should be that desired on the vapor pressures. As the most reproducible results are usually most readily obtained at pressures of about one atmosphere, this is probably the best pressure to choose. The inert gases chosen should preferably be two whose equation of state constants are as different as possible.

(2) Values of p and the corresponding mol fractions for various values of the pressure, with one of the two inert gases used above, at certain temperatures covering the temperature range desired. These data need not be as accurate as the one atmosphere data.

(3) Values of the molal volume of the liquid or solid and the solubility of the inert gases used over the temperature range. Since the corrections by Eq. (2) are in general small the accuracy required is not very great.

At a given pressure and temperature the value of the actual vapor pressure p_1 will be independent of the inert gas used except for a small term in y. Since, in the type of experiments used, p is in general much larger than p_1, $(p-p_1)$ will vary negligibly from gas to gas at a given total pressure. Consequently by applying Eqs. (1) and (2) to each inert gas at constant temperature and total pressure p, the term in $(p-p_1)$ can be eliminated and there results, using the asterisk to designate values for the second inert gas,—

$$\frac{2.303(RT)^2}{p}\left(\log\frac{x_1^*}{x_1}-\log\frac{y^*}{y}\right) = (A_{01}^{1/2}-A_{02}^{1/2})^2x_2^2-(A_{01}^{1/2}-A_{02}^{*1/2})^2x_2^{*2} + R\frac{(c_1^{1/2}-c_2^{1/2})^2x_2^2-(c_1^{1/2}-c_2^{*1/2})^2x_2^{*2}}{T^2}. \quad (3)$$

In almost every case for gas current vapor pressures the difference between x_2^{*2} and x_2^2 will be small enough so that an average value may be used, so that the right hand side of the equation reduces to

$$\left[2A_{01}^{1/2}(A_{02}^{*1/2}-A_{02}^{1/2})+A_{02}-A^*_{02} + R\frac{2c_1^{1/2}(c_2^{*1/2}-c_2^{1/2})+c_2-c_2^*}{T^2}\right]x_2^2 \equiv Mx_2^2. \quad (4)$$

If values of M are plotted against $1/T^2$ a line will result whose intercept determines A_{01} and whose slope determines c_1. The normal experimental errors are such that the slope will seldom be determined very accurately and usually must be readjusted after using the variable pressure data.

Using the approximate values of A_{01} and c_1 found in this way, a value of B_{01} can be found which when Eqs. (1) and (2) are applied to the variable pressure data gives a value of p_0 which is constant within the experimental error. This is simpler to carry out than might appear, since in most cases an assumed approximate value of p_0, one for example based on the one-atmosphere value of px_1, applied to the right hand side of Eq. (2) and the same value used for p_1 on the right hand side of Eq. (1) will be sufficient for all except the final calculations. Unless the values of A_{01} and c_1 are right the values of B_{01} thus calculated will show a trend with temperature. This means that the line drawn on the M, $1/T^2$ plot from Eq. (4) must be readjusted in slope and new values of A_{01} and c_1 used. Usually, if B_{01} decreases as the temperature increases, it indicates that the c_1 value and consequently the slope chosen was too large. This readjustment is carried on till the deviations on the M, $1/T^2$ plot are not excessive, and a satisfactory constancy of B_{01} with temperature is obtained.

The final values of A_{01}, B_{01} and c_1 are then used in Eq. (1) to correct the more accurate one-atmosphere px_1 observations to give p_1. Normal vapor pressures may then be calculated by Eq. (2).

Constants thus obtained can probably be justifiably used in determining also the pressure, volume, temperature relationships for the vapor of the liquid at low pressures by the use of the Beattie-Bridgeman equation of state for low pressures by means of either of the two simple forms

$$V/n_1 = RT/p + [B_{01} - A_{01}/RT - c_1/T^3]$$
$$p = n_1RT/V + [B_{01} - A_{01}/RT - c_1/T^3]n_1^2RT/V^2.$$

The accuracy of these corrections on the perfect gas law will probably be considerably less when so used than when the constants are used to calculate corrections on the same type of data from which they were determined; but in many cases, such as the case of iodine which we are about to discuss, there is little possibility of being able to determine the equation of state with equal accuracy by direct measurement.

The Vapor Pressure of Iodine

Practically the only substance for which suitable data exist at present for the application of this method is iodine. Braune and Strassman[8] have recently published data from which the variation of px_1 from one atmosphere to about forty atmospheres with carbon dioxide could be calculated over a temperature range of about 60°C. They also give observations at one atmosphere with both carbon dioxide and hydrogen as inert gases. Due to a lack of smoothness of their vapor pressures as a function of temperature, the best

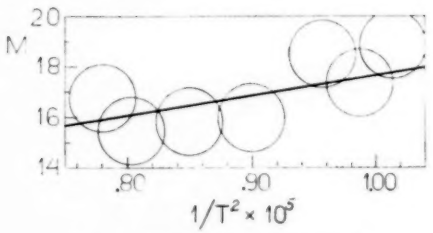

Fig. 1. Plot of the function M.

results for normal vapor pressures will probably be obtained by correcting the results by Baxter, Hickey and Holmes[9] and of Baxter and Grose,[10] who used air as an inert gas. However, the use of a different gas for the determination of vapor pressures from either of those used in determining the constants introduces a certain possibility of error in the calculation of the corrections, as will be discussed more fully in the next section.

Unfortunately Braune and Strassman do not give their direct experimental quantities but calculated values, which fact necessitates a reversal of their calculations to get their experimental variables p and x_1. At ten temperatures they give values at one atmosphere for hydrogen and carbon dioxide. Three of these are sufficiently inconsistent to fall off the plot in Fig. 1 where the quantity of M of Eq. (4) is plotted against $1/T^2$. The radii

[9] Baxter, Hickey and Holmes, J. Am. Chem. Soc. **29**, 127 (1907).
[10] Baxter and Grose, J. Am. Chem. Soc. **37**, 1061 (1915).

of the circles correspond to an accumulated error of 0.2 percent on their combined results for hydrogen and carbon dioxide at the temperature in question. The line corresponds to the final values of A_{01} and c_1 chosen after obtaining a proper value of B_{01} as described above.

The values of the constants found for iodine are given in Table. I. Carbon dioxide and hydrogen were assumed insoluble in solid iodine and the Inter-

TABLE I. *Values of the constants for iodine vapor in the Beattie-Bridgeman equation of state for low pressures.*

Constant	A_{01}	B_{01}	c_1	R	Molecular weight
Value	17.0	0.325	4000×10^4	0.08206	253.864

Units: atmospheres, liters, moles, $T=t°C+273.13$.

national Critical Tables[11] values for the density of solid iodine were used. The deviations in percent of the observed values of the quantity px_1/p_1 in the high pressure observations with carbon dioxide from the values calculated by Eq. (1) using these constants are shown in Fig. 2. The circles

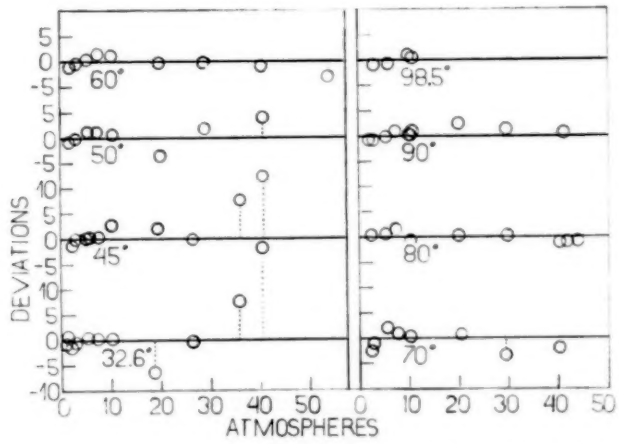

Fig. 2. Deviations: px_1/p_1, observed minus calculated, in percent of the calculated.

correspond to the claimed experimental accuracy of the measurements. It can be seen that the deviations are not systematic from temperature to temperature except possibly for the low temperature higher pressure points where we might well expect the low pressure equation we are using to be insufficient, since the magnitude of the correction in that region is about the size of p_1 itself.

Since the one-atmosphere observations are probably more exact than those at higher pressures, the one-atmosphere points were used to calculate the normal vapor pressure of iodine with the aid of the above constants. For the purpose of showing the importance of the corrections, the values of

[11] International Critical Tables **3**, 21 (1928).

px_1 as well as the normal vapor pressures obtained from them are given in Table II, together with the best values from the high pressure carbon dioxide experiments.

TABLE II. *Values of px_1 at one atmosphere, and normal vapor pressures of iodine in millimeters of mercury.*

	Values of px_1		Normal vapor pressures			
$t°C$	px_1 CO_2	px_1 H_2	High p CO_2	p_0 CO_2 1 atm.	p_0 H_2 1 atm.	Average of last two
32.6	0.5778	0.5608	0.5687	0.5642	0.5635	0.5639
45.0	1.504	1.467	1.482	1.473	1.474	1.4735
50.0	2.243	2.185	2.207	2.199	2.196	2.197
60.0	4.281	4.191	4.244	4.206	4.211	4.208
70.0	8.254	8.095	8.185	8.124	8.131	8.127
80.0	15.23	14.96	15.14	15.01	15.02	15.015
85.0	21.20	20.82	—	20.91	20.90	20.90

A comparison of the px_1 values with the corrected values of the vapor pressure shows that the corrections are quite significant and bring about a much better agreement between the two sets of data.

The normal vapor pressures calculated from the data of Braune and Strassman are found not to be smooth when their logarithms are plotted against $1/T$. This is however true of the original observations.

The measurements of Baxter and his co-workers, [9,10] who used dry, carbon-dioxide-free air as an inert gas in the gas current method at one atmosphere, appear very accurate. The vapor pressures are a smooth function of the temperature.

As already noted, [5] the algebra of combination of constants is such that a mixture such as air may without any inconsistency be treated as a pure substance in applying Eq. (1). The values of the constants used for air were those determined for air [6,7] when dry and free from carbon dioxide, rather than by using the method of combination of constants for its constituent gases.

Using these values for the constants of air and the values in Table I for iodine the normal vapor pressure was found for various temperatures from the data given by Baxter and Grose. The data are represented by the equation given by them:

$$\log p' = 9.7522 - \frac{2863.54}{273 + t° - 19} \tag{5}$$

where p' is equal to px_1 under the supposition that the volumes of dry air may be correctly calculated from the measured volumes of moist air by the use of the customary procedure involving the ideal gas laws.[12]

With the aid of Eqs. (1) and (2) and the assumption that air is insoluble in solid iodine, values of the normal vapor pressure of iodine were calculated

[12] This supposition we shall here accept. The method which we are here applying may in principle be applied to improve the usual procedure, but at present the necessary constants for water are lacking.

from the data in Eq. (8) for even values of the temperature. These are given in Table III together with the corresponding values of px_1.

TABLE III. *Values of px_1 and of the normal vapor pressure of iodine in millimeters of mercury according to the measurements of Baxter and collaborators by the gas current method.*

t°C	0	10	20	30	40	50
px_1	0.03009	0.0804	0.2001	0.4670	1.0287	2.1511
p_0	0.02981	0.0798	0.1988	0.4643	1.0235	2.1419

t°C	60	70	80	90	100
px_1	4.292	8.206	15.092	26.79	46.04
p_0	4.276	8.179	15.047	26.71	45.89

The effect of the correction is decidedly important, compared with the precision obtainable in the best measurements by the gas current method. The normal vapor pressures of Braune and Strassman are in approximate agreement with those given in Table III; the average deviation being about 1.6 percent when only their one-atmosphere observations with hydrogen and carbon dioxide are considered.

Giauque[13] has recently smoothed the data of Baxter and his co-workers with the aid of spectroscopic and specific heat data. Application of his method to the above corrected results of Baxter gives an equation which fits the data well within the probable experimental accuracy. This equation is

$$\log p_{0\text{atm.}} = -\frac{3512.8}{273.1 + t^\circ} - 2.013 \log (273.1 + t^\circ) + 13.3740. \qquad (6)$$

Due to the theoretical advantages of this form, the vapor pressures calculated from it are probably more accurate than those given in Table III using Baxter's empirical smoothing function. The coefficient of the reciprocal absolute temperature is the only constant which differs numerically from that used by Giauque, as the other constants were derived by him from specific heat and spectroscopic data. The agreement of Eq. (6) with the data is slightly better than the empirical Eq. (5).

A Comparison of Methods of Calculation

It can readily be seen from the results for iodine that even in the case of observations taken at as low a pressure as one atmosphere, the commonly used assumption that vapor pressures can be calculated from the gas current method data directly from the pressure mol fraction product may be seriously in error. Since it is known[1,2] that the assumption of this form of Dalton's law includes most of the perfect gas law we should expect deviations with real gases, but even where this is recognized it is often assumed that at such low pressures these deviations are within the experimental error.

The second usual method of calculation, that using Gibb's form of Dalton's law, makes no restriction on the equation of state of pure gases though it involves the rule of additivity of pressures[14] for mixtures. Both the addi-

[13] Giauque, J. Am. Chem. Soc. **53**, 507 (1931).

[14] Gibbs "Collected Works of J. Willard Gibbs", **1**, 155, Longmans Green and Co., 1906 and 1928.

tivity of pressure rule and the Gibbs form of Dalton's law have experimentally been shown to be inexact for real gas mixtures. In many of the cases studied this form of Dalton's law is superior to the previously mentioned form but this is not necessarily the case, as can be readily seen from studying the identity

$$\frac{px_1/p_1}{(n_{1m}/V_m)/(n_1/V)} = \frac{pV_m/(n_{1m}+n_{2m})}{p_1V_1/n_1} \tag{7}$$

where p, V_m, n_{1m}, n_{2m}, are the pressure, volume, moles of substances 1 and 2 in a gas mixture, and p_1, V_1, and n_1 are corresponding values for pure gas 1 in equilibrium with it through a semipermeable membrane. The deviation of $(n_{1m}/V_m)/(n_1/V)$ from unity gives the deviation from the Gibbs-Dalton law, as the deviation of px_1/p_1 gives it for the simple rule. By using a low pressure form of the Beattie-Bridgeman equation of state corresponding to Eq. (1) we have for the pressure volume product of a pure gas

$$pV/n_1 = RT + [B_{01} - A_{01}/RT - c_1/T^3]p \tag{8}$$

and for a gas mixture, where the summations are over the various components of the mixture,

$$pV_m/\sum n_1 = RT + [\sum x_1B_{01} - (\sum x_1A_{01}^{1/2})^2/RT - (\sum x_1c_1^{1/2})^2/T^3]p. \tag{9}$$

By combining the last three equations we can calculate the deviations from the Gibbs-Dalton law from that known for the simple rule. This has been done for two temperatures in air, hydrogen and carbon dioxide as inert gases with iodine, and the deviations from both methods of computation are given in Table IV. It can be seen that the deviations from the Gibbs-Dalton law are of the same order of magnitude and in general slightly greater than for the simple rule. This is what we would expect from Eq. (7) in the case of small mol fractions where p is much greater than p_1 as is the usual case in gas current vapor pressure determinations. It can be seen that this method is equally unsatisfactory for our purpose as the first.

TABLE IV. *Deviations in percent in the calculation of the vapor pressure of iodine in various gases by simple methods.*

Inert gas	Air		H_2		CO_2	
t°C	25	75	25	75	25	75
$100\left(\frac{n_{1m}}{V_m}-\frac{n_1}{V_1}\right)\Big/\frac{n_1}{V_1}$	0.45	0.07	−0.77	−0.73	2.95	1.57
$100\frac{px_1-p_1}{p_1}$	0.42	0.13	−0.71	−0.61	2.43	1.31

A graphical method of extrapolating the px_1 values for various values of p to the limit where p equals p_0 is much more likely to be accurate than the two previous methods. There is here however an uncertainty due to the dependance of the corrections on the two variables p and x, which vary simul-

taneously, and a rapid change of x with p takes place at values of p (near p_0) below the values of p usually obtaining in gas current vapor pressures. This may be shown as follows.

The customary method for obtaining constants or virial coefficients in an equation of state for gas mixtures is to use a function of the form

$$b_m = b_1x_1^2 + 2b_{12}x_1x_2 + b_2x_2^2 \tag{10}$$

where b_1 and b_2 are the constants of the gases 1 and 2 and x_1 and x_2 the corresponding mol fractions with b_{12} an interaction constant. Any such method applied to the β term, the only one of importance at low pressures, or to its component constants, in a virial expansion in p of the equation of state of a mixture

$$pV_m/\sum n_1 = RT + \beta_m p + \cdots \tag{11}$$

gives for the equilibrium pressure calculation a relation of the form

$$2.303RT \log (p_1/px_1) = \beta_1(p - p_1) + F_{12}px_2^2 \tag{12}$$

where β_1 is the β of equation (11) for the pure gas 1 and F_{12} is a function of the constants of gas 1, gas 2 and their interaction constants. Such a function gives

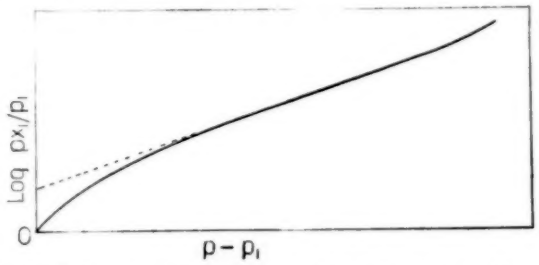

Fig. 3. A typical plot, showing failure of purely empirical treatment.

a curve of the general form shown in Fig. 3 for the value of log px_1/p_1 against $(p-p_1)$. It can be seen that when the mol fraction x_1 becomes very small at higher pressures the results are nearly linear but the straight line does not pass through the origin. Since the experimental results in gas current vapor pressure determinations are usually in this higher pressure range, direct extrapolation is likely to introduce appreciable errors. This error will usually be negligible when the vapor pressure is less than two or three millimeters.

Extrapolation of values of the quantity $(n_{1m}/V_m)/(n_1/V_1)$ against the density of the inert gas offers theoretical advantage in that the extrapolation should be very nearly linear. However, this method offers in practice uncertainties equal to or greater than that using pressures, as the usual experimentally measured quantities are the total pressure and the mol fractions. In order to obtain mols per unit volume assumptions must be made about the equation of state of gas mixtures, and further assumptions about the equation of state of the pure vapor are necessary to convert the resulting extrapolated density into a vapor pressure. Due to the considerable possibility of substantial systematic errors from this source it seems advisable to use only methods involving pressures and mol fractions.

The method we are using in this paper uses the functions for β_1 and F_{12} in Eq. (12) taken from the Beattie-Bridgeman equation of state for pure gases and gas mixtures, which has received considerable experimental justification. It seems possible that certain small changes in the equation may have to be introduced in certain instances. These, however, are not of such a nature as to invalidate our method but only to limit the amount of extrapolation that is possible of the results from the conditions of temperature and of inert gases used in determining the constants to quite different conditions.

In the case where the available variable pressure data are with the same inert gas as the best one-atmosphere data from which the normal vapor pressures are to be determined, the method we are using can be regarded as simply a means of determining the amount of curvature at low pressures illustrated in Fig. 3 for a single isotherm and of smoothing the corrections as a function of the temperature. Under these circumstances we should expect very precise results in the determination of the normal vapor pressures over the experimental temperature range.

Where the corrections are applied to data with a different inert gas from either of those used in determining the constants of the equation, the possibilities of error are multiplied. We have applied it above in this way in the case of the iodine-air work of Baxter and his collaborators. In a case such as this, particularly as the corrections are intermediate between those for the two inert gases used in determining the constants, it seems entirely probable that the corrections are of the right order of magnitude though probably of considerably less accuracy than those calculated for carbon dioxide. The uncertainties of the method as a whole appear in any event to be less than those of the other methods used for the calculation of normal vapor pressures from data of the gas current method.

VARIATION OF THE VAPOR VISCOSITIES OF NORMAL AND ISOPENTANE WITH PRESSURE BY THE ROTATING CYLINDER METHOD

RALPH K. DAY
RESEARCH ASSISTANT, AMERICAN PETROLEUM INSTITUTE

(Received January 26, 1932)

ABSTRACT

The vapor viscosities of both normal and isopentane have been measured by the rotating cylinder method at different pressures ranging from a few millimeters pressure to the saturated vapor pressures all at 25°C. Twenty-eight measurements are given for normal pentane and thirty-seven for isopentane with an estimated accuracy of 0.2 percent. The plotted values for each compound shows a linear decrease with pressure, the slopes being nearly the same. The results may be expressed for normal pentane $\eta=(677.2-0.0084p)\times10^{-7}$ poises; for isopentane $\eta=(696.5-0.0077p)\times10^{-7}$ poises where p is the pressure of the vapor in millimeters of mercury.

A CAREFUL study of the vapor viscosities of the isomeric pentanes was thought worth while first, because there exist three isomers alike in molecular weight and chemical composition but whose molecular shapes are probably quite different. The relation of their viscosities might shed light on their relative molecular structure. A second reason for selecting the pentanes was because their boiling points are near room temperature and the rotating cylinder method can be used to measure the viscosity of a vapor close to its saturated vapor pressure. The large distance between the cylinder (6 mm) allows measurements of high accuracy in spite of minute condensations which might occur. The capillary tube method is out of the question for viscosity measurements on saturated vapors. Unfortunately, the symmetrical tertiary pentane has not yet been obtained so this paper presents only measurements on normal and isopentane vapors.

APPARATUS

The rotating cylinder apparatus used for this work has been constructed with the greatest care to give results of the highest accuracy. It was copied in its essential features and dimensions after that used by Harrington,[1] Gilchrist[2] and others at the University of Chicago and is described in detail in my Doctor's thesis[3] on the same subject. Fig. 1 shows a schematic drawing of the apparatus to scale. A brief description, mostly of unique details, follows:

The cylinders themselves were specially cast of homogenous bronze and carefully turned to the correct dimensions. Twenty-one measurements taken at different points on the rotating cylinder's inner surface with micrometer

[1] Harrington, Phys. Rev. **8**, 738 (1916).
[2] Gilchrist, Phys. Rev. **1**, 124 (1913).
[3] Thesis, R. K. Day, Calif. Institute of Technology (1930).

calipers gave a maximum difference between any two readings of 0.006 mm which is 5 parts in 100,000 of its diameter and one part in 1000 of the distance between the cylinders. The inner cylinder was only 1/64 inch in thickness with six ribs 1/4 inch wide and 1/64 inch thick turned in place and the

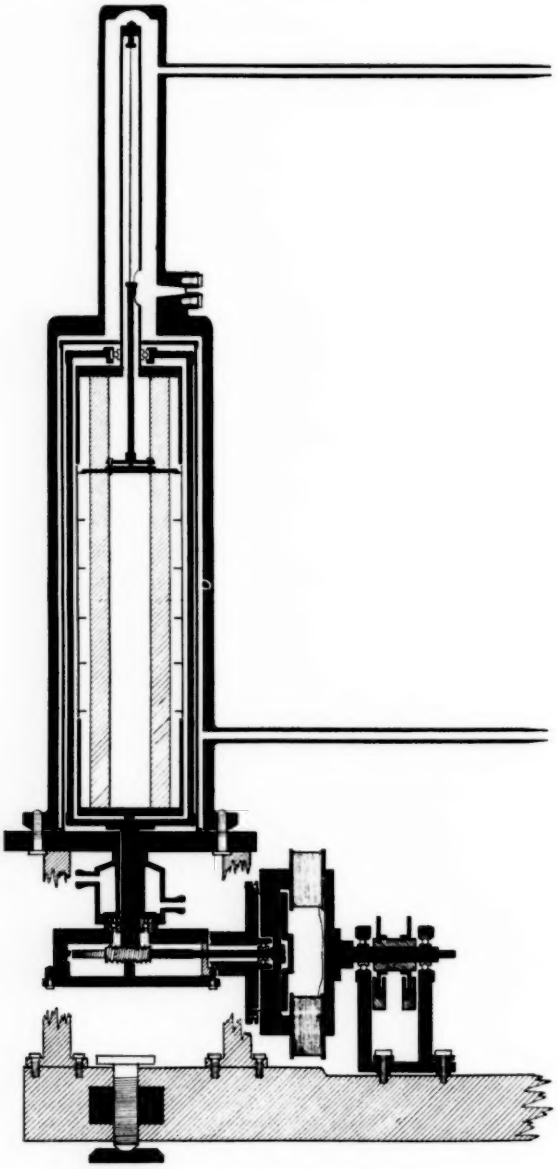

Fig. 1. Diagram of apparatus.

diameter could not be measured directly because of its frailty. It was supported in a heavy iron cylinder while the last lathe cut was taken on the inside and the accuracy of its surface should be the same as that of the outer cylinder.

The outer, rotating cylinder is supported by dry ball bearings and is driven from the bottom by means of an agate worm driving a cast iron gear all inside the large vacuum-tight apparatus housing. To transmit the rotation of the driving motor through the brass housing to the worm inside, a magnetic coupling was used which proved entirely satisfactory.

The driving motor was copied after a motor designed by H. Benioff for use in driving the chronograph drums on the Seismographs in use at the Carnegie Seismograph Station in Pasadena. Its speed-controlling mechanism and driving unit is a vacuum-tube-driven steel reed mounted on a heavy base in a thermostated box. This constant frequency is amplified with vacuum tubes and relays and fed directly to the motor poles. The accuracy and reproducibility of the speed of this motor is better than one part in 10,000 as estimated by Mr. Benioff.

The inner cylinder is suspended on an uncoiled watch hair spring of cross section 0.018″×0.058″ kindly supplied by the Elgin Watch Co. The high sensitivity obtained necessarily gives a long natural period of oscillation to the inner cylinder system (198.3 sec.), so that a special damping device had to be resorted to. Three coils of insulated wire were symmetrically placed around the outside of the brass apparatus casing and connected in a closed delta to a three phase a.c. source. A current of 0.2 amps. in this "three-phase induction motor" was sufficient to twist the suspended cylinder inside. By means of a reversing switch next to the observer, the suspended cylinder could be brought to rest rapidly near its equilibrium deflected position so that readings could be taken rapidly in spite of the long natural period of oscillation of the cylinder.

Summary of Apparatus Dimensions and Constants

Length of rotating cylinder	18–1/4 in.
Inside diameter of rotating cylinder	4.690 in.
Length of each guard cylinder	3–7/8 in.
Length of suspended cylinder	10–1/16 in.
Outside diameter of suspended cylinder	4.218 in.
Weight of suspended cylinder	478.5 gms.
Length of suspension	9–3/8 in.
Natural period of suspended cylinder	198–1/3 sec.
Speed of rotation of outer cylinder (approx.)	1 rev./min.
Worm gear drive ratio	120:1

Deflection Measurements

A maple meterstick whose center was 214 cm from the mirror attached to the inner cylinder was viewed through the mirror with a telescope to measure the deflections. Tenths of a millimeter could be estimated and deflections for the pentanes of about 170 mm gave an accuracy of reading of 1 part in 1700. The scale was accurately perpendicular to a line from its center to the center of the mirror so that angular deflections were computed as:—

$$2\theta = \tan^{-1}\frac{x_l}{d} + \tan^{-1}\frac{x_r}{d}$$

where x_l and x_r are the readings measured from the center of the meterstick obtained when the outer cylinder was run successively in opposite directions and d is the mirror-to-scale distance.

Temperature Control

This work was done in a sub-basement room thermostated at 25° Centigrade. The temperature never varied from this by more than 0.2°C while mostly it was within 0.1°C of 25°C. Using Sutherland's formula and the values of the constants given in the International Critical Tables a change in the viscosity of air of 1 part in 1800; of pentane of 1 part in 1300 is caused by 0.2°C change of temperature.

Purity of Materials

The two isomers of pentane which occur in natural gasoline were obtained in a very pure state. Normal pentane was obtained directly from the Eastman Research Laboratory. Isopentane was found unavailable in a sufficiently pure state, so that our Chemistry Department synthesized it for this work.

To test the purity of these compounds, the freezing points and the range of temperature over which freezing takes place were measured. This physical property was chosen first, because freezing point is a good criterion of purity and second, because the difference in the freezing points of the two isomers is 28°C whereas much smaller differences occur in most of the other physical properties. The pentane was put into a small poorly evacuated Dewar flask, which was immersed in a larger Dewar containing liquid air. By means of a copper constantan thermocouple and with constant stirring, cooling curves were plotted giving the following results compared with the values of the freezing points given in the International Critical Tables.

Table I

	Freezing Point		Difference	Range of Freezing
	Meas.	Int. Crit. Tables		
N. pentane	−132.8°C	−131.5°C	1.3°C	1.0°C
Isopentane	−161.1°C	−159.7°C	1.4°C	0.3°C

The nearly constant difference between the measured values and the I. C. T. values of the freezing points is likely due to error in the thermocouple used.

Method of Handling Pentanes

Normal and Isopentane have boiling points of about 36° and 28°C, respectively. This low boiling point and the small amount of material available necessitated handling the the material only when it was frozen or very cold in liquid air. Pentane while frozen in a Pyrex flask, is sealed onto a side arm

connected through an all metal stopcock B to the apparatus. (Vacuum system diagram Fig. 2). Since the vapor pressure of the pentanes is negligible at liquid air temperatures, the whole apparatus could now be evacuated with the metal stopcock open. The evacuating apparatus system was now isolated from the pumps with a large stopcock A, the metal stopcock separating the pentane from the apparatus closed and the liquid air removed from the pentane. As the pentane warmed up successive quantities could be intro-

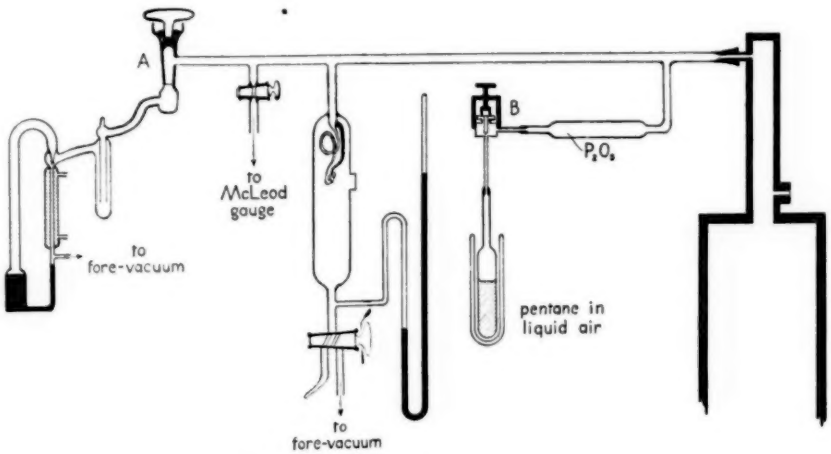

Fig. 2. Diagram of vacuum system.

duced into the apparatus by opening the metal stopcock for an appropriate time. The all metal stopcock was resorted to because no greased stopcock or even phosphoric acid or sealing wax stopcock was satisfactory. Pressure measurements were made by means of a small Pyrex Bourdon spiral which was used as a zero instrument balanced against a closed tube mercury manometer. The purpose of this was to protect the metal apparatus from the mercury vapor. Pressure measurements were accurate to 1 mm of mercury.

Procedure of Measurements

Three independent sets of measurements were made on each isomer. A set of measurements consisted of a succession of deflection readings taken with increasing amounts of pentane introduced into the apparatus by successive openings of the metal stopcock. At the top pressure liquid air was put around the pentane flask and by again successively opening the metal stopcock measurements were taken at decreasing pressures until most of the pentane was again frozen out of the apparatus. At each pressure step the outer cylinder was rotated first in one direction and then in the other following a definite time schedule. Using the three-phase damping control, the deflected cylinder was quickly brought nearly to rest after which from three to six small swings about the deflected position were observed. Computation of the viscosity was made from the mean of small oscillations at the deflected positions in the two directions of rotation of the outer cylinder without

measuring the zero equilibrium position of the inner cylinder, according to the formula given.

All measurements were made at the same temperature, with the same motor speed and with the same suspension. The standard formula of the rotating cylinder apparatus[1] is:—

$$\eta = \frac{\pi\theta I(b^2 - a^2)}{a^2b^2P^2\Omega l}$$

θ is the deflected angle; I, the moment of inertia of the suspended system; b and a the diameters of the outer and inner cylinders; P, the natural period of the suspended system; Ω, the angular velocity of the rotating cylinder and,

TABLE II. *Normal pentane:* $\theta_{air}=4°.5101$.

	No.	p	t	x_1	x_r	$\eta\times10^7$
	7	9	15	17.101	14.964	676.98
	8	17	35	17.108	14.970	677.22
	9	25	55	17.099	14.982	677.30
	10	50	75	17.090	14.976	677.02
Run	11	102	95	17.060	14.965	676.29
No. 1.	12	50	115	17.088	14.967	676.77
	13	25	135	17.086	14.979	676.98
	14	17	155	17.079	14.979	676.81
	15	9	175	17.085	14.971	676.77
	16	0.38	14	17.016	14.908	673.97
	17	1.3	34	17.100	14.994	677.51
	18	4	54	17.118	15.014	678.40
	19	9	74	17.124	15.021	678.64
Run	20	51	94	17.090	14.976	677.02
No. 2.	21	146	114	17.069	14.954	676.08
	22	250	134	17.048	14.935	675.23
	23	200	154	17.058	14.940	675.56
	24	100	174	17.078	14.948	676.16
	25	25	199	17.083	14.958	676.49
	26	8	224	17.089	14.956	676.57
	27	200	28	17.080	14.912	675.43
	28	300	63	17.058	14.882	674.34
	29	402	88	17.046	14.881	674.05
Run	30	479	210	17.041	14.855	673.40
No. 3.	31	484	230	17.038	14.852	673.28
	32	455	300	17.030	14.853	673.12
	33	350	345	17.062	14.856	673.85
	34	250	365	17.080	14.890	674.97

l, the length of the inner cylinder. No attempt has been made at absolute measurements because the viscosity of air has been so accurately determined by Harrington that his value can be used for calibration purposes. All of the instrument constants now fall out and the relative viscosity equation reduces to:—

$$\frac{\eta}{\eta \text{ air}} = \frac{\theta}{\theta \text{ air}}$$

Harrington's value for the viscosity of air at 23°C can be corrected to 25°C using Professor Milliken's[4] formula as follows:—

[4] R. A. Millikan:—Ann. d. Physik **41**, 759 (1913).

$$\eta_{air}(\text{at } 23°\text{C}) = 1822.6 \times 10^{-7} \text{ (Harrington)}$$

$$\eta_T = \eta_{23°} + 0.000000493\,(T - 23) \text{ (Millikan)}$$

This gives

$$\eta_{air}(\text{at } 25°\text{C}) = 1832.5 \times 10^{-7} \text{ C.G.S. units.}$$

Results

In the following table "p" is the pressure in mm of mercury, "t" is the time in minutes between the reading and the time when the apparatus was shut off from the pumps. The calibration deflection on air at 25°C is taken as the mean of a calibration run immediately before and after the readings taken.

Table III. *Isopentane:* $\theta_{air} = 4°.5081$.

	No.	p	t	x_1	x_r	$\eta \times 10^7$
	1	2.9	30	17.860	15.124	696.60
	2	11.5	55	17.882	15.142	697.45
	3	25	75	17.868	15.141	697.12
Run	4	50	95	17.858	15.118	696.42
No. 1.	5	101	115	17.850	15.112	696.13
	6	50	135	17.855	15.111	696.22
	7	24	155	17.857	15.108	696.20
	8	11.5	175	17.855	15.110	696.20
	9	2.1	15	17.939	15.031	696.30
	10	11.5	35	17.954	15.035	696.70
	11	25	55	17.946	15.030	696.42
	12	50	75	17.942	15.023	696.20
	13	101	95	17.924	15.012	695.59
Run	14	150	115	17.916	15.008	695.32
No. 2.	15	201	135	17.906	14.997	694.89
	16	251.5	155	17.901	14.990	694.63
	17	350	175	17.894	14.983	694.34
	18	250	195	17.905	14.976	694.42
	19	150	215	17.919	14.991	695.05
	20	50	235	17.934	14.999	695.53
	21	25	15	17.966	15.003	696.28
	22	50	35	17.960	14.995	595.98
	23	100	55	17.951	14.986	695.60
	24	202	75	17.929	14.975	694.91
	25	300	95	17.910	14.956	694.12
	26	350	115	17.908	14.946	693.87
Run	27	400	135	17.903	14.945	693.74
No. 3	28	450	165	17.901	14.932	693.43
	29	497	195	17.871	14.924	692.62
	30	550	235	17.868	14.919	692.46
	31	650	448	17.874	14.911	692.41
	32	658	525	17.867	14.903	692.10
	33	600	570	17.855	14.902	691.84
	34	553	595	17.875	14.906	692.33
	35	400	625	17.892	14.915	692.89
	36	300	645	17.897	14.934	693.40
	37	149	665	17.931	14.957	694.58

Discussion of the Accuracy of the Results

The biggest concern with the accuracy of the results was the retention of the purity of the compounds. There was no way in which the liquid pen-

tane could gather impurities through introduction into the apparatus because it was always frozen out again and the small residue in the apparatus discarded. Since the apparatus had been previously evacuated no volatile substances could remain to distill back into the pentane flask with the pentane. However, while the vapor is in the apparatus small air leaks and residual gases coming off of the apparatus walls which were invariably present could easily contaminate the vapor. 0.2 percent of air was introduced intentionally and the result showed no detectable difference of viscosity so that

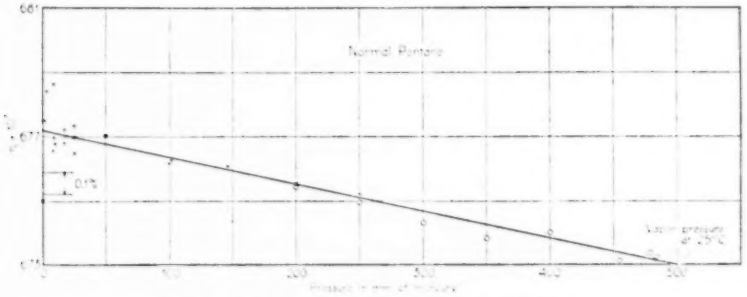

Fig. 3. Viscosity of normal pentane vapor.

small air impurities were probably of small concern. The method of taking measurements by first successively increasing the pressure and then freezing out again was used because any errors due to foreign gas impurities would increase with the time after the apparatus was shut off from the pumps. For this reason this period of time is included in the table of results and it will be noticed that measurements of the viscosity at the same pressure and after different lengths of time after evacuation check within the limits of error.

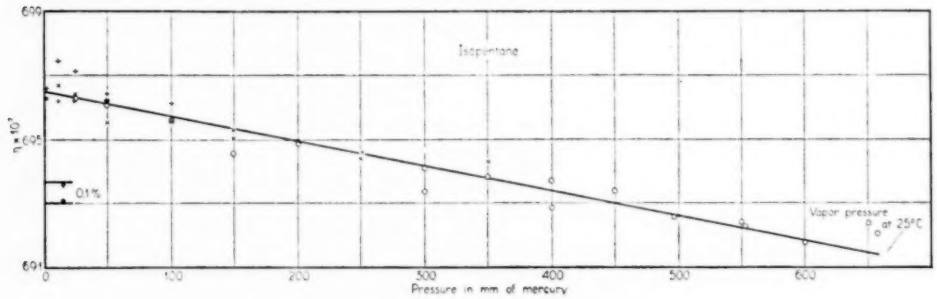

Fig. 4. Viscosity of isopentane vapor.

Errors due to fluctuations in the speed of the motor or any imperfections in the apparatus are believed to be within the accuracy of reading because of the consistencies of the readings taken on air and hydrogen and an accurate check of the air: hydrogen viscosity ratio.

There was a slight drift ("nachwirkung") in the suspension but all readings including calibration readings were made following the same time schedule so that any small error incurred due to suspension drift was balanced out. Readings were taken on air with and without a 2:1 ratio gear inserted

between the motor and apparatus. The angular deflections were accurately in the same 2:1 ratio which is assurance that the apparatus works free of turbulance according to the theoretical formula used.

It will be noticed that the largest deviations between readings at the same pressure are in the very low and very high pressure ranges. These are the difficult pressures at which to work. The apparatus, stopcocks, rubber gaskets and other materials inside it tend to absorb the pentane. These absorptions produce the most trouble in the very high and very low pressure ranges. At low pressures, small quantities of the evaporating or condensing vapors are large percentages of the total vapor present and at pressures near saturation very large amounts of pentane, even puddles, may accumulate. The evidence for this latter is in the large changes in pressures on standing which are noticed at the high pressures. During these changes of pressure, the deflections were always uncertain which was probably due to currents of vapor blowing on the inner cylinder before equilibrium was attained.

The error of reading the deflection is about 1 part in 1700 or 0.06 percent. The error due to temperature irregularities is about 0.08 percent. Other errors are probably much less so that the over-all error may be estimated to be 0.2 percent. It will be noticed that nearly all the observed points are within 0.1 percent of the straight line drawn through them and *all* the observations taken in the three runs on each compound have been included.

Previous Measurements

A search of the literature revealed only one reference to viscosity measurements of these compounds; namely that of Rappenecker for isopentane at 100°C and 215.5°C using the capillary tube method. His values are:—

$$\eta = 885.1 \times 10^{-7} \text{ at } 100°C$$
$$\eta = 1164 \times 10^{-7} \text{ at } 212.5°C$$

Sutherland's constant (C) = 500.

Extrapolating his values with the help of his Sutherland constant and Sutherland's equation, one obtains $\eta = 691 \times 10^{-7}$ at 25°C. He used "Kahlbaum" isopentane and worked at a few centimeters of mercury above atmospheric pressure. This extrapolated value agrees with my high pressure value of 691.5×10^{-7} at 650 mm.

Discussion of Results

After all the measurements were made and plotted, it was quite apparent that a straight line fits best the points of each curve. It will be further noticed that the slopes of the lines on the two isomers are of the same order of magnitude (ratio:85/76). The difference between the absolute values at any pressure is about 2.8 percent, the isopentane having the greater viscosity.

These values are as would be expected. The isopentane being a more compact molecule would be expected to have a smaller collision area and con-

[5] Rappenecker, Zeits. f. physik. Chem. **72**, 695 (1910).

sequently a longer mean free path and larger viscosity. The decrease in viscosity with increase of pressure on each compound seems to indicate an increase in the effective size of the molecule with increase of pressure. This fact and the apparent linear relationship need careful theoretical consideration. Kinetic theory interpretation of this decrease of viscosity with increase of pressure and of the apparent linear relationship may lead to some interesting and new knowledge of the behavior of gaseous vapors from the molecular viewpoint.

Summary

The vapor viscosities of both normal and isopentane have been accurately measured at 25°C by the rotating cylinder method at different pressures ranging from a few millimeters pressure to the saturated vapor pressures. Each isomer shows a linear decrease of viscosity with increasing pressure, the total decrease in each case being slightly less than one percent. The difference between the absolute values of the viscosities of the isomers at any pressure is about 2.8 percent, the isopentane having the higher viscosity. This result is consistent with what would be expected from the organic chemist's structural picture of these isomers.

The results may be expressed as follows:—

$$\text{normal pentane: } \eta_{25} \times 10^7 = 677.2 - 0.0084p \text{ poises}$$
$$\text{isopentane: } \eta_{25} \times 10^7 = 696.5 - 0.0077p \text{ poises}$$

where p = pressure of the vapor in millimeters of mercury.

This research was carried out as a project of the American Petroleum Institute who paid for the apparatus and most of the work. The work was carried out under the personal direction of Professor R. A. Millikan and much assistance was given by Mr. Wm. Bleakney. Credit is due Mr. Julius Pearson and his instrument shop for help in design and the high accuracy of construction of the apparatus.

THE DIELECTRIC CONSTANT OF DICHLORETHANE, DIBROMETHANE, CHLOROBROMETHANE, AND DIACETYL; AND THE PHENOMENON OF FREE ROTATION

By C. T. Zahn
Paris, France

(Received February 29, 1932)

Abstract

Measurements of the dielectric constant of dichlorethane, recently published by the author, indicate a considerable temperature variation of the electric moment, as had been predicted by various authors, because of the possibility of free rotation around the simple C–C bond, and a simultaneous dipole interaction between the two mutually rotating parts of the molecule. From these results one would predict a similar temperature variation in other symmetrically substituted ethanes containing large characteristic moments such as those of C–Cl, C–Br, and C=O. In order to verify this supposition the present measurements were made on dibromethane, chlorobromethane, and diacetyl. The results prove to be similar to those for dichlorethane, and hence consistent with the predicted behavior. The variations in electric moment ($\times 10^{18}$ c.g.s.e.s.u.) for the four substances are as follows: dichlorethane (previously published) 1.12–1.54 in the temperature interval 305–544°K; dibromethane 0.94–1.10 in the interval 339–436°K; chlorobromethane 1.09–1.28 in the interval 339–436°K; and diacetyl 1.25–1.48 in the interval 329–504°K. A short discussion is finally given of the evidence for free rotation in ethane obtained from the specific heats of ethylene and ethane, and the possible importance of the hydrogen atoms in the interactions of the dichlorethane molecule.

IN A recent article the author published the results of measurements of the dielectric constant of dichlorethane vapor.[1] This work was undertaken as an attempt to clarify certain discrepancies[2] in the literature concerning the existence of a temperature variation of electric moment associated with the possibility of free rotation. The above mentioned results seem to be consistent with a considerable temperature variation of the electric moment of dichlorethane; namely, from 1.1 to 1.5×10^{-18} c.g.s.e.s.u. in the temperature interval from 32 to 270°C. Dichlorethane represents one of the simplest compounds in which one would expect to find evidence for the existence of free rotation. From the chemical point of view this evidence is furnished by the inability to find chemically separable *cis* and *trans* isomers. From the physical point of view measurements of the electric moment furnish the possibility of obtaining a more complete knowledge of the degree of freedom of this predicted motion around the C–C bond. As is now well known these latter measurements indicate that this rotation is neither truly free nor is it com-

[1] C. T. Zahn, Phys. Rev. **38**, 521 (1931).

[2] L. Meyer, Zeits. f. physik. Chem. (B) **8**, 27 (1930); R. Sänger, Phys. Zeits. **32**, 21 (1931); L. Meyer, Phys. Zeits. **32**, 260 (1931); R. Sänger, Phys Zeits. **32**, 414 (1931).

pletely bound by the interaction between the two mutually rotating parts of the molecule. When the interaction is more and more appreciable one would expect the moment to correspond more and more nearly to that of the molecule in the position of minimum potential energy, that is presumably the *trans* position for which $\mu = 0$. One would also expect a temperature variation of the moment as a consequence of the temperature variation of the average state of rotational vibration about the position of minimum potential energy. As has been pointed out the measurements of the dielectric constant of dichlorethane are consistent with a considerable interaction between the two rotating CH_2Cl groups.

In general in molecules where the two rotating parts themselves have large characteristic moments one would expect the interaction to be great; and if the interaction is not too great one would expect to observe a variation of electric moment in the observable temperature range. For dichlorethane this is the case, since the moment of C–Cl is about 1.5×10^{-18} c.g.s.e.s.u. (that of C–H is only 0.4×10^{-18}). In order, therefore, to make a more general study

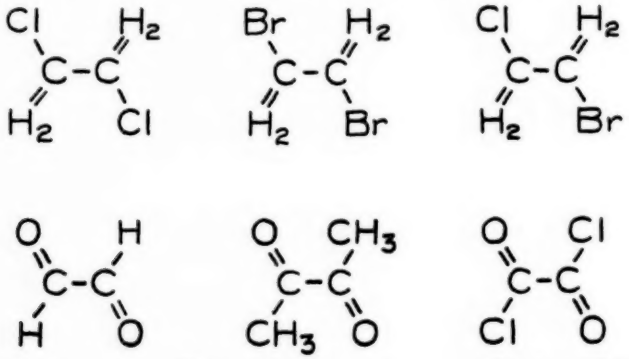

Fig. 1. dichlorethane — dibromethane — chlorobromethane
glyoxal — diacetyl — oxalyl chloride

of this phenomenon, the present measurements have been made on three other substituted ethanes containing large characteristic dipoles. First dibromethane was studied; here the change in moment from C–Cl to C–Br is negligibly small, but the size and mass of the groups are changed considerably. Next chlorobromethane was studied; here a dissymmetry in the substituents is introduced.

Since the electric moment characteristic of the C=O bond is about 2.3×10^{-18} it would be interesting to have information concerning the electric moment of glyoxal. This substance, however, presents experimental difficulties, both physical and chemical, which make such a study impracticable. The next simplest choice was diacetyl, where the hydrogen atoms of glyoxal are replaced by methyl groups. Here one has the possibility of a study of the interaction chiefly between two C=O bonds. Another compound of considerable interest in this connection is oxalyl chloride, where there exist the simultaneous interactions between the four bonds of the two types C=O and C–Cl. This compound seems to be adaptable to such a study for practical

reasons, but so far it has unfortunately been impossible to extend the present measurements to include it.

The results of the present measurements on the three substances dibromethane, chlorobromethane, and diacetyl are included in Table I. In the second column P represents the molecular polarization in cm^3 calculated with the small correction for pressure obtained by the use of the van der Waals constants as previously described.[3] Since this latter correction is small an extremely accurate knowledge of the van der Waals constants is not important. For dibromethane the constants were obtained directly from Landolt-Börnstein; for chlorobromethane an average was taken of the values for

TABLE I.

T°K	P	max. p (cm Hg)	$\mu\times10^{18}$ c.g.s.e.s.u
	Dibromethane		$A=29.7$
339.12	45.81	10	0.94
367.78	46.08	19	0.99
405.22	45.95	20	1.03
435.76	46.82	20	1.10
	Chlorobromethane		$A=26.4$
338.93	48.12	10	1.09
368.22	48.11	21	1.14
405.28	50.87	21	1.20
435.60	49.66	21	1.28
	Diacetyl		$A=22.7$
328.80	51.90	10	1.25
358.74	51.42	20	1.29
391.04	50.75	20	1.33
425.95	49.92	20	1.37
461.09	49.65	20	1.42
504.33	49.66	20	1.48

dichlor- and dibromethane; and for diacetyl the value for acetic anhydride was used. The errors in such approximations are probably only a small fraction of a percent.

In the last column of Table I are given the values of the electric moment μ, calculated from the value of the constant part of the electric polarization A. As has been customary the value of A is obtained from refractivity data from Landolt-Börnstein, with ten percent added to the refractivity for the Na D line, as an estimate for the contribution of the infrared terms. Such approximations are, of course, somewhat questionable; but in most cases the orientational contribution to the electric polarization is larger than the constant part, and the resulting possible error in the electric moment is probably of the order of one percent. For example, in the case of dichlorethane the

[3] C. T. Zahn, Phys. Rev. **35**, 848 (1930).

value of A thus obtained is in good agreement with Hitchcock's value of P measured from the solid state. In the case of chlorobromethane an average was taken for the refractivity as in the case of the van der Waals constants. The refractivity of diacetyl does not seem to have been measured. In this case a value was computed on the usual assumption of additivity of refractivity from characteristic constants as tabulated by Errera.[4]

The computed values of electric moment are plotted against the temperature in Fig. 2. For the sake of comparison the previous values for dichlorethane[1] are included in this graph. As was to be expected all four substances

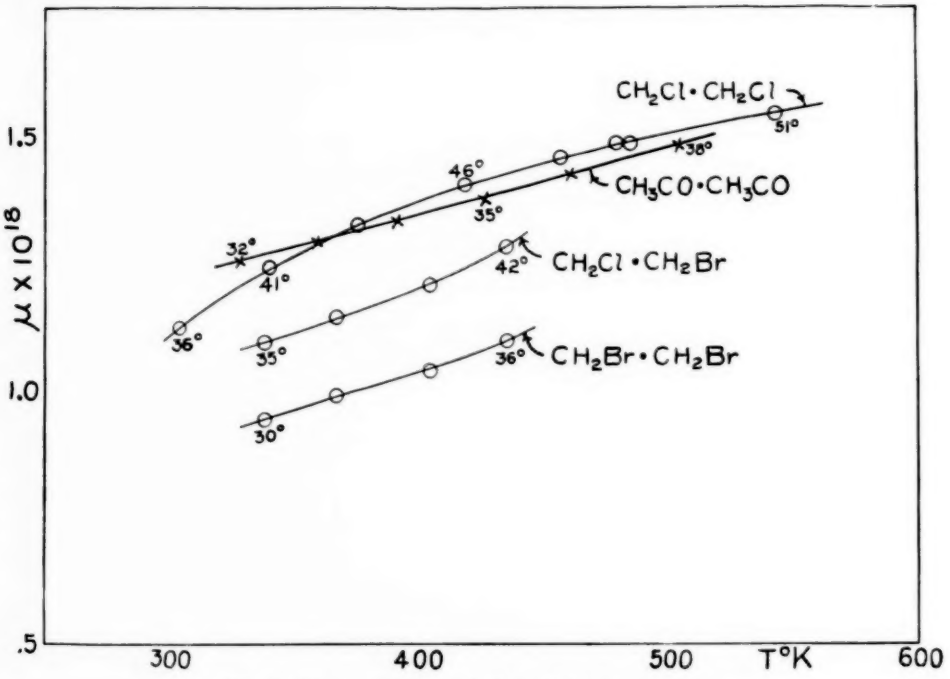

Fig. 2. Temperature variation of electric moment.

show a definite increase in electric moment with the temperature. Dibromethane has a considerably smaller moment than dichlorethane even though the characteristic moments of the two are approximately the same. This indicates that there is a stronger interaction between the brominated groups than between the chlorinated groups. The curve for chlorobromethane lies approximately midway between those for dichlor- and dibromethane, as might be expected. Finally it is seen that the electric moment of diacetyl is nearly the same as that of dichlorethane. This fact indicates a greater interaction between the CH_3CO groups than between the CH_2Cl groups, consistent with the larger characteristic moment of the former.

In order to obtain a rough idea of the magnitude of the "amplitude" of the rotational vibrations consistent with these data, one can calculate the

[4] J. Errera, Polarisation Diélectrique, p. 75.

rotational angle corresponding to the observed electric moment. For this calculation the method of Eucken and Meyer was used; namely, to assume tetrahedral valence symmetry and fixed characteristic electric moments for each chemical bond, as customarily obtained from previously measured electric moments. The following values for these characteristic moments were used;[5] C=O, -2.3; C–Cl, -1.5; C–Br, -1.5; and C–H, $+0.4\times10^{-18}$ c.g.s.e.s.u. In this manner the corresponding angles were calculated for the extreme temperatures and a few intermediate temperatures, as indicated in Fig. 2.

In Table II these results are tabulated together with the computed values of moment corresponding to free rotation.

TABLE II.

Substance	T°K	$\mu\times10^{18}$ c.g.s.e.s.u. (observed)	$\theta°$	$\mu\times10^{18}$ c.g.s.e.s.u. (free rotation)
$CH_2Cl\cdot CH_2Cl$	305–554	1.12–1.54	36–51	2.54
$CH_2Br\cdot CH_2Br$	339–436	0.94–1.10	30–36	2.54
$CH_2Br\cdot CH_2Cl$	339–436	1.09–1.28	35–42	2.54
$CH_3CO\cdot CH_3CO$	329–504	1.25–1.48	32–38	3.20

DISCUSSION OF EVIDENCE FOR FREE ROTATION

Concerning the possibility of rotation around the C–C bond in dichlorethane, and similar compounds, there are three principal cases which may be differentiated according to the nature of the potential energy of the molecule as a function of the rotational angle.

(1) If there is no interaction between the two CH_2Cl groups other than the axial part associated with the C–C bond, and the potential energy does not therefore vary with the angle of rotation, one would expect completely free rotation. In this case the expected moment should be about $2^{1/2}\times1.8 = 2.5\times10^{-18}$. This case is obviously excluded by the experimental values of the moment.

(2a) If the potential energy function has one very deep trough sharply defined within narrow limits of rotational angle, this angle would remain practically constant. Hence the molecule should be practically rigid and the measured moment should not vary with the temperature.

(2b) If the potential energy function has several such deep troughs there should exist several corresponding chemically inseparable rotational isomers, in proportions depending on their energies and the temperature. If the electric moments of these isomers are not all equal, there should result a temperature variation in the observed electric moment corresponding to the change in distribution of the isomers.

(3) If the minima of cases (2a) or (2b) are not sharply defined the molecule should no longer have effectively rigid angular positions. There should

[5] See H. Sack, Ergebnisse der Exakter Naturwissenschaften **8**, 338–340 (1929).

then exist either nonuniform rotations or rotational vibrations, and the instantaneous electric moment should vary over a wide range. Since these motions should vary with the temperature, there should also exist a corresponding temperature variation of the observed electric moment.

Now the experimental data on the dielectric constant of dichlorethane, and the other similar substituted ethanes, indicate a considerable temperature variation of the electric moment (unless one is willing to admit, what seems improbable, that the constant part of the molecular polarization A is practically doubled on passing through the infrared region, and that the variations in P are due to experimental error.) This would exclude case (2a) and leave cases (2b) and (3) as the only remaining possibilities, since these two cases alone would give a temperature variation in observed electric moment. These two cases differ only in the degree in which the minima of potential energy are sharply defined.

From the observed variation in electric moment it is at present impossible to obtain an accurate knowledge of the nature of the potential energy function for a molecule of the type of dichlorethane; but it will probably be of interest to consider the matter from the point of view of other existing experimental evidence. It has been customary in the case of this molecule to consider the interaction as chiefly a simple dipole repulsion between the two C–Cl bonds, with a somewhat arbitrary location of the dipoles on the C–Cl axes; but it should be remembered that the range of distribution of the electricity giving rise to the dipoles is of the same order of magnitude as the atomic separations. Hence the nature of the interactions might be quite different from what one would predict from the simple dipole model. For example, there might be pronounced minima in potential energy corresponding to positions other than the *trans* position, caused by the presence of the hydrogen atoms. On the simple dipole model this seems improbable because of the small moment characteristic of the C–H bond and the relatively small size of the hydrogen atoms, but there seems to be evidence from the specific heats of ethylene and ethane pointing toward the possibility of a considerable interaction due to the C–H bonds.

In fact the experimental values of the specific heats of ethylene and ethane indicate that in ethane the rotation is not free but that there exists a rotational vibration for which the specific heat has practically reached the classical value even below room temperature. Ebert[6] has discussed these specific heat data and arrived at the conclusion that it would be impossible to distinguish between the cases of free rotation and of rotational vibration completely set up in the classical sense. Wagner[7] has recently pointed out a miscount of the number of additional degrees of freedom to be associated with free rotation. As Ebert has pointed out, the experimental data indicate two extra degrees of internal freedom for ethane. Now free rotation can actually account for only one, instead of two as Ebert supposed, and it should there-

[6] L. Ebert, Leipziger Vorträge (1929) p. 74.

[7] Carl Wagner, Zeits. f. phys. Chem. (B) **14**, 166 (1931).

fore be possible to distinguish between the above two cases. It therefore seems necessary to assume that there is a rotational vibration completely set up, and that the other extra degree of internal freedom corresponds to the potential energy associated with these vibrations. From symmetry considerations one would expect three equivalent positions of minimum potential energy spaced 120° apart. If these rotational vibrations actually exist in ethane, one would expect that for temperatures considerably above room temperature the motion would tend toward that of free rotation (that is, as kT becomes greater than the binding energy). Then the number of extra degrees of internal freedom should tend from two to one. As a matter of fact it is interesting to note that Ebert's curves do show a slight decrease in this extra internal heat, although this slight decrease is probably not outside the limits of precision of the calculations. The specific heat data at the basis of these calculations all correspond to temperatures below room temperature. It would therefore be very interesting to have experimental data on the specific heats of ethylene and ethane above room temperature, in order to determine whether the predicted decrease in internal specific heat at high temperatures actually exists or not. The temperature at which the transition toward free rotation becomes marked would give an idea of the energy of binding associated with the supposed rotational vibrations.

In the above case of ethane, if the three minima are sharply defined, one could consider ethane as a mixture of three equivalent rotational isomers. Such rotational isomers have been supposed to exist in the case of more complicated substituted ethanes of the type of tartaric acid, because of the temperature variation of the optical activity of such compounds.[8] In such cases, however, it is not surprising that there should exist pronounced multiple minima, since there is the possibility of considerably close proximity of the larger OH and COOH groups. Even here it does not seem possible from the experimental evidence to ascertain accurately the sharpness of definition of such rotational isomers.

Returning again to the case of dichlorethane, and similar substituted ethanes, the values of the electric moment indicate that the effect of the Cl repulsion is predominant, since the average calculated configuration is only 30–50° from the *trans* position. This does not, however, exclude the possibility of the existence of small proportions of sharply defined isomers, of considerably larger angle, and caused by the interactions associated with the H atoms.

Other evidence concerning the structure of dichlorethane has been furnished by the x-ray diffraction experiments of Debye[9] and the electron diffraction experiments of Wierl.[10] Both these experiments indicate separations of the two Cl atoms approximately corresponding to the *trans* position of the Cl atoms, consistent with the above interpretations of the dielectric constant data. Since a considerable departure from the *trans* position does

[8] See R. Lucas, Ann. de Physique, X• Série, **9**, 26 (1928).
[9] P. Debye, Phys. Zeits. **31**, 142 (1930).
[10] R. Wierl, Phys. Zeits. **31**, 366 (1930).

not change the Cl separation very much, these experiments are not very sensitive for determining the average departures in configuration of the molecule from the *trans* position.

In order to make the experimental evidence in favor of the interpretations discussed in this article more conclusive, there are several interesting experiments which it is perhaps worth while to mention here in the hope that they may be performed in the near future. First it would be very valuable to have experimental data on the specific heat of ethylene and ethane above room temperature, in connection with the discussions of Ebert and Wagner and the nature of the interactions between C–H bonds. Then similar experiments on dichlorethane would serve as further evidence as regards internal rotation. Finally it would be very interesting to have molecular ray experiments on the electric moment of dichlorethane.

The experiments here reported were performed at the Palmer Physical Laboratory, Princeton University.

THE PROBLEM OF THE ROTATING MAGNET

By John Zeleny and Leigh Page
Yale University

(Received March 5, 1932)

Abstract

The analysis of the problem of the rotating magnet previously given by us is extended. For the special cases considered, it is shown that, while the magnet exerts equal and opposite torques on the portions of the current circuit inside and outside of the magnet, neither of these portions individually exerts any torque on the magnet itself. The analysis is applied also to the case where the magnet is replaced by a solenoid, and to the case where the magnet is replaced by a brass cylinder and the flux through it is supplied by an outside coaxial solenoid. Pietenpol and Westerfield's recent criticism of our previous analysis is shown to be mistaken, and the fault in their method of attacking the problem is found to lie in the application of Ampère's original law to the interaction between two current elements where in general it is invalid.

SOME years ago[1] we deduced a simple formula for the torque acting on a cylindrical magnet when a current is flowing through it. In the special cases considered, we found that the current as a whole produces no effect upon the magnet and that the whole torque results from the action of the magnet on the current inside of it and in portions rigidly connected with it. In a recent paper,[2] Pietenpol and Westerfield make a claim that the forces which we found to be the effective ones could in reality only produce stresses within the magnet, and they give an alternate derivation, leading to the same formula as ours, based on the action of the Amperian currents in the magnet upon the portion of the circuit not rigidly connected with the magnet.

Kimball,[3] too, does not altogether like our method of treating the problem because he feels it involves "the necessity of thinking of the flux carried by the magnet as fixed to outer space" and prefers to think of this flux as fixed rather to the moving magnet. We must emphasize the fact that we made no assumption as to the lines of force being fixed in space and that as far as the electromagnetic equations are concerned the question of motion of lines of force is quite irrelevant. Kimball obtains a result identical with ours by finding the torque exerted by the field of the magnet upon the part of the circuit not attached to the magnet and then invoking the law of action and reaction to get the torque acting on the magnet itself. This method of solving the problem is perfectly correct although it is less direct and more laborious than ours. We shall show that in reality the field of the magnet exerts equal and opposite torques on the part of the circuit inside the magnet and on that outside. We computed directly the torque acting on the current inside the

[1] J. Zeleny and L. Page, Phys. Rev. **24,** 544 (1924).
[2] W. B. Pietenpol and E. C. Westerfield, Phys. Rev. **38,** 2280 (1931).
[3] A. L. Kimball, Jr., Phys. Rev. **28,** 1302 (1926).

magnet, which is the one we wished to find, and did not think it necessary even to point out the existence of the equal and opposite torque acting on the rest of the circuit, which Kimball has obtained.

Now lines of force are after all only devices by whose aid the consideration of many problems is simplified. Engineers and others have found it helpful to think of the lines as attached to one object or another, but this concept must be used guardedly as it may lead to erroneous results in cases where the object is moving relative to the observer. It is necessary to appeal to the electromagnetic equations to determine whether the use of this concept leads to the correct result in any specific case. We must point out that the idea of lines of force being attached to any object is not a part of current electromagnetic theory and that according to this theory a field of force is simply a vector quantity, the representation of whose sense and magnitude by lines is purely a matter of convenience.

The law which we used for the force acting on a current element located in a magnetic field is one of the fundamental laws of electrodynamics. In its application, no limitation is imposed by the place of origin of the magnetic field in question nor by the state of motion that an investigator feels inclined to attribute to lines of force. Nor is there any necessity of considering the reactions involved. There can be no question as to the applicability of the law to the problem of a rotating magnet. By its use we have obtained in a simple manner a solution of the problem which agrees with experiment and which is admittedly correct.

In our previous paper we were not interested in showing all of the reactions between the elements of the mechanism under consideration. We did prove theoretically and verify experimentally that when the current enters the side of the magnet through an arm which for some distance is rigidly connected with the magnet, the effect on this arm is opposite to that upon the current inside of the magnet, and stated that if this arm were infinitely long the total torque on the moving system would vanish. This indicates that the reaction in the fixed part of the system as ordinarily used is equal and opposite to that in the movable portion.

We propose now to offer a further analysis of these reactions and to show that, although the magnet exerts equal and opposite torques on the part of the circuit inside the magnet and on that outside, nevertheless, neither of the above portions of the circuit exerts any torque whatsoever on the magnet itself. We already proved in our previous paper that the circuit as a whole exerts no torque on the magnet. We shall also point out what we believe to be the weakness in Pietenpol and Westerfield's method of attack and the error in some of their contentions.

Analysis of the Interactions

The magnet, in the form of a right circular cylinder, is supposed to be free to rotate about its axis of symmetry. The current enters the magnet through a side arm or apron of zero or finite length and passes out along the axis of the magnet. The circuit, therefore, consists of two parts, a fixed exterior

portion and a movable interior portion which lies in the side arm and magnet. Mechanically the system consists of two parts, the fixed exterior portion of the circuit on the one hand, and the magnet and the interior portion of the circuit on the other hand.

Assuming, as is permissible in the usual set up, that the interactions between the interior and exterior portions of the circuit are relatively negligible, there remain to be considered the four torques involving the magnet, as follows:

1. the torque $L_{i_1}{}^m$ exerted by the magnet on the current inside the magnet and side arm or apron,

2. the torque $L_{i_0}{}^m$ exerted by the magnet on the fixed, exterior portion of the circuit,

3. the torque $L_m{}^{i_1}$ exerted by the current in the interior portion of the circuit on the magnet, and

4. the torque $L_m{}^{i_0}$ exerted by the current in the exterior portion of the circuit on the magnet.

We have, then, for the total torque on the movable part of the system

$$L_1 = L_{i_1}{}^m + L_m{}^{i_1} + L_m{}^{i_0}, \tag{1}$$

and for the torque on the fixed part of the system

$$L_2 = L_{i_0}{}^m. \tag{2}$$

As is well known the torque exerted by one complete circuit on a second complete circuit is equal and opposite to the torque exerted by the second on the first. So as the magnet consists of an assemblage of Amperian circuits, $L_{i_1}{}^m + L_{i_0}{}^m$ must be equal and opposite in sign to $L_m{}^{i_1} + L_m{}^{i_0}$, that is

$$L_1 + L_2 = 0. \tag{3}$$

Our interest is in the torque L_1 on the movable part of the system, that is, on the magnet and side arm or apron. Two cases have been considered: (I) the case where the current in both the exterior and the interior portions of the circuit is in the form of a sheet or sheets symmetrically disposed with respect to the axis of the magnet but the magnet is not necessarily symmetrically magnetized, and (II) the case where the magnet is magnetized symmetrically with respect to its axis but neither the external nor the internal currents are necessarily symmetrical.

In our previous treatment of this problem we showed that in both cases

$$\left.\begin{aligned} L_{i_1}{}^m &= iN/2\pi, \\ L_m{}^{i_1} + L_m{}^{i_0} &= 0. \end{aligned}\right\} \tag{4}$$

We now proceed to show that the torque $L_m{}^{i_0}$ exerted by the current in the exterior portion of the circuit on the magnet vanishes in each of the two cases under consideration.

Case I. The exterior portion of the circuit is in the form of a symmetrical current sheet.

Then symmetry requires that the field due to the exterior portion of the circuit shall be symmetrical with respect to the axis of the magnet. Now the torque on any arbitrarily selected Amperian circuit in the magnet is equal to the increase in the magnetic flux through it per unit change in angular position. But on account of the symmetry of the field, rotation of the magnet about its axis produces no change in the flux through any one of the Amperian circuits which constitute it. Hence the torque on each Amperian circuit and consequently on the magnet as a whole is zero, irrespective of whether the magnet be symmetrically magnetized or not.

Case II. The magnet is symmetrically magnetized.

Consider Amperian circuits lying in a torus of small cross-section whose axis coincides with that of the magnet. On account of the symmetrical magnetization the density, orientation, etc. of the Amperian circuits is uniform all the way around the torus. So if the torus is rotated about the axis of the magnet the total flux through all of the Amperian circuits contained in it due to an asymmetric field of the exterior portion of the current circuit remains unchanged, the increased flux through those which move from weaker to stronger parts of the field being balanced by the decreased flux through those which move from stronger to weaker parts of the field. Therefore the net torque on the torus under consideration and hence on the whole magnet vanishes. This conclusion may be reached as well by another line of argument. In a symmetrically magnetized cylinder, the Amperian currents actually present may be replaced by circular currents lying in planes perpendicular to the axis of the cylinder with centers on this axis. This is done by cancelling coincident equal and opposite currents, as indicated in a previous paper,[4] and cannot affect the torque on the magnet. But the force on a current element due to a magnetic field of any character whatsoever is at right angles to the current element. Therefore the force on each element of one of the circular currents under consideration passes through the axis of rotation, and consequently gives rise to no torque.

Having shown that $L_m{}^{i_0}$ vanishes, it follows from (4) that $L_m{}^{i_1}$ vanishes also in each of the two cases under discussion.

The net result of our analysis of the problem, then, is as follows. The magnet exerts a torque $iN/2\pi$ on the interior portion of the current and an equal and opposite torque on the exterior portion of the current. *Neither portion of the current exerts any torque on the magnet.* In so far as the movable and fixed portions of the mechanical system are concerned, the law of action and reaction holds as regards the torques involved; that is to say, the torque exerted on the movable part of the system is equal and opposite to that exerted on the fixed part of the system, as is shown by Kimball's analysis taken in conjunction with our original analysis of the problem. On the other hand, the

[4] J. Zeleny and L. Page, Phys. Rev. **27**, 427 (1926).

present analysis shows that the reactions between the various *electrically separable elements* constituting the system do not consist of equal and opposite torques.

Pietenpol and Westerfield point out that the torque on an asymmetrically magnetized cylinder is given by the formula $iN/2\pi$ provided only the exterior portion of the current circuit is symmetrical, no matter whether the interior portion is symmetrically distributed or not. That this contention is correct can be shown very simply by the following considerations. Suppose first that the current is symmetrical inside the magnet as well as outside. Then the torque L_2 exerted by the magnet on the exterior portion of the circuit is $-iN/2\pi$. Now redistribute the current inside the magnet so that it is no longer symmetrical, leaving the exterior current unchanged. Evidently the torque L_2 on the exterior portion of the circuit has not been affected, since the magnetic field to which it is due has remained unaltered. Therefore the equal and opposite torque L_1 on the suspended system is unchanged. Obviously a similar argument does not apply to the case where the interior portion of the current circuit remains unchanged and the symmetrical exterior current is replaced by an asymmetrical distribution.

Criticism of Pietenpol and Westerfield's Analysis

Pietenpol and Westerfield calculate directly, by means of Ampère's original formula for the force between two current elements, the torque ${L_m}^{i_0}$ exerted on the Amperian currents of which the magnet is composed by the exterior portion of the circuit. To this torque, for which they find the value $iN/2\pi$, they ascribe the rotation of the magnet. Furthermore they conclude that, as the formula which they use states that the force acts along the line joining two current elements, the torque ${L_{i_0}}^m$ exerted on the exterior portion of the circuit by the magnet must be equal and opposite to ${L_m}^{i_0}$. Their results, in conjunction with Eqs. (4), imply that

$$ {L_{i_1}}^m = {L_m}^{i_0} = -{L_m}^{i_1} = -{L_{i_2}}^m = iN/2\pi \tag{5} $$

and that therefore the torque ${L_{i_1}}^m$ exerted by the magnet on the interior portion of the circuit, to which we ascribed the rotation, is annulled by the equal and opposite torque ${L_m}^{i_1}$ exerted by the interior portion of the circuit on the magnet. This contention is in direct contradiction with the results of our analysis given above, which shows definitely that the torques ${L_m}^{i_0}$ and ${L_m}^{i_1}$ are individually zero for each of the two cases under consideration. We shall show now wherein the fault of Pietenpol and Westerfield's analysis lies, and point out the reason why they obtain the correct expression for the resultant torque although they ascribe it to the wrong interaction.

Ampère's original law of force, as Pietenpol and Westerfield remark, gives correct results for the force between one complete circuit and a part of another circuit, but generally gives incorrect results for the force between two partial circuits. Pietenpol and Westerfield seem to be under the impression that they have applied the law only to the first of these two cases. However, in calculating the torque exerted by the exterior circuit on an Amperian cur-

rent in the magnet (reference 2 p. 2282) they have applied the law to the interaction of one partial circuit on another and in this way have been led to an unjustified conclusion. On the other hand, their method, if used to find the torque exerted by the Amperian currents in the magnet on the exterior portion of the circuit, involves only the force between one complete circuit and a current element of another and hence leads to correct results. Now, the formula which they have used requires the first of these torques to be equal and opposite to the second, so their misapplication of the law to the calculation of the first torque gives them the correct value of the second, and therefore the proper expression for the resultant torque on the magnet.

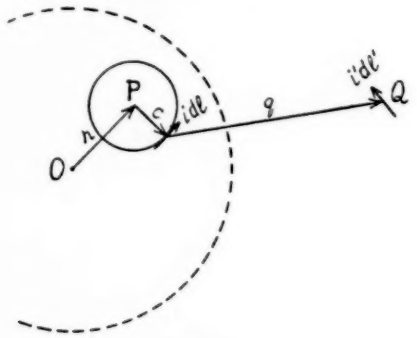

Fig. 1.

To be more specific, Pietenpol and Westerfield consider the torque on a circular Amperian current at P in the magnet (Fig. 1) about some point O on its axis of symmetry due to the force exerted by the current element $i'dl'$ of the exterior portion of the circuit. If $\mathbf{r}$ is the position vector of the center P of the Amperian current and $\mathbf{c}$ is its vector radius, the torque on the element idl of this circuit is (in their notation)

$$d^2\mathbf{T} = (\mathbf{r} + \mathbf{c}) \times d^2\mathbf{F}, \tag{6}$$

where $d^2\mathbf{F}$ is the force between idl and $i'dl'$. The total torque on the Amperian circuit due to $i'dl'$ is, then, the sum of the two torques

$$\begin{aligned} d\mathbf{T}_1 &= \mathbf{r} \times d\mathbf{F}, \\ d\mathbf{T}_2 &= \int \mathbf{c} \times d^2\mathbf{F}. \end{aligned} \tag{7}$$

Now $d\mathbf{T}_1$ involves only the force on the *entire* Amperian current due to $i'dl'$, but $d\mathbf{T}_2$ involves the force exerted by $i'dl'$ on *each element* idl of the Amperian current. So the original form of Ampère's law of force is applicable to the calculation of $d\mathbf{T}_1$ but cannot be expected to give correct results when used to calculate $d\mathbf{T}_2$.

On the other hand the torque on $i'dl'$ is

$$d\mathbf{T} = \int (\mathbf{r} + \mathbf{c} + \mathbf{q}) \times (-d^2\mathbf{F}) = \int (\mathbf{r} + \mathbf{c}) \times (-d^2\mathbf{F}), \tag{8}$$

a formula which gives correct results since only the force exerted on $i'dl'$ by the *entire* Amperian current is involved, and it makes no difference into what components this force may be resolved since they are all applied at the one point Q. Hence Pietenpol and Westerfield's method gives correct results for the torque on the exterior portion of the circuit, but incorrect results for the torque exerted by the exterior portion of the circuit on the magnet. Only because their law of force requires the two torques to be equal and opposite, does their faulty calculation of the second lead to the correct value of the first.

The electromagnetic equations are quite specific in their assertion that the force exerted by one current element on another is not in general along the line joining the two. For a current element is nothing more than a small charge de moving in one direction with velocity $\boldsymbol{v}_1$ coincident with an equal charge of opposite sign moving in the opposite direction with velocity $\boldsymbol{v}_2$. Provided the velocities of the charges are small compared to the velocity of light, Maxwell's equations lead unambiguously to the formula

$$d\boldsymbol{H} = \frac{de(\boldsymbol{v}_1 - \boldsymbol{v}_2) \times \boldsymbol{r}}{r^3} = \frac{id\boldsymbol{l} \times \boldsymbol{r}}{r^3} \tag{9}$$

for the field strength at distance $\boldsymbol{r}$, where $id\boldsymbol{l}$ is the equivalent current element. The force exerted by this field on a second current element $i'd\boldsymbol{l}'$ is given by the force equation

$$d^2\boldsymbol{F} = i'd\boldsymbol{l}' \times d\boldsymbol{H} = \frac{ii'}{r^3}\{d\boldsymbol{l}' \times (d\boldsymbol{l} \times \boldsymbol{r})\}, \tag{10}$$

which shows that the force is not in general along the line joining the two current elements. In fact, if we have two parallel current elements (or two charges moving parallel to each other) as in Fig. 2, the magnetic forces which

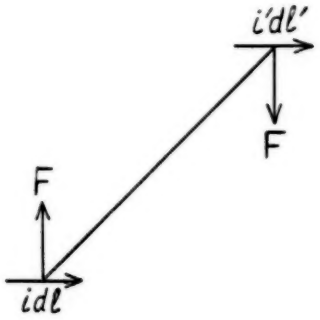

Fig. 2.

they exert on each other do not lie in the line joining them, but give rise to the unbalanced couple $\boldsymbol{F}$, $\boldsymbol{F}$. So while Ampère's original formula leads to correct results for the force between a complete circuit and a part of a circuit, it cannot be applied to calculate the force between two incomplete circuits.

Magnet Replaced by Solenoid

We showed in our previous paper[1] that when the suspended magnet is replaced by a brass cylinder surrounded by a fixed coaxial solenoid, the cylinder experiences a torque given by the same formula as that applicable to the cylindrical magnet, the flux N being now due to the solenoid. Pietenpol and Westerfield agree with our explanation in this case, but they assert that were the solenoid rigidly attached to the rotating cylinder the torque experienced by the movable system would be the result of the action of the fixed exterior portion of the circuit upon the solenoid. Our preceding analysis shows clearly, however, that no steady magnetic field can exert a torque on the solenoid about its axis of symmetry, provided the latter is wound so as to have no axial component of current. Accordingly, whether the solenoid be attached to the cylinder or not, it exerts equal and opposite torques on the cylinder and the exterior portion of the circuit, but neither of these parts exerts any torque on the solenoid itself.

Unfortunately it is not possible to distinguish experimentally between the two points of view, since it is only possible to measure the resultant torque on each part of the mechanical system and in either case the torque experienced by the solenoid if it were made movable would be zero since it is implied in Pietenpol and Westerfield's reasoning that the cylinder exerts on the solenoid a torque equal and opposite to that which the latter exerts on it.

It would be possible, too, to dispense with the brass cylinder and pass the current in the exterior fixed portion of the circuit through a suspended solenoid, having the current enter the solenoid at a point on its periphery and leave at a point on its axis. In such a case the torque on the solenoid would be exerted by the field produced by its circular turns on the axial and radial portions of the current.

Conclusion

We have in the device of the rotating magnet, then, an example where angular actions and reactions are equal and opposite in so far as the *mechanically separable* parts of the mechanism are concerned but do not occur in equal and opposite pairs between the *electrically separable* elements of the system. In systems of this type the magnetic field serves merely as a non-mechanical intermediary for the interaction of the two parts of the mechanical system.

PRINCIPLES OF A NEW PORTABLE ELECTROMETER

By Ross Gunn
Naval Research Laboratory, Washington, D.C.

(Received February 15, 1932)

Abstract

The fundamental principles underlying the operation of a new type of portable electrometer are given. It is shown that the periodic transfer of static charges from one capacity system to another and back again can be used to produce an alternating electrical potential. The alternating potential is amplified by means of vacuum tubes, then rectified by a synchronous commutator and the resulting direct current used to operate an indicator. Instruments with capacities of about 25 cm and sensitivities of 10^4 microamperes per volt have been built to operate with portable microammeters used as the indicator. The period of the device depends on the indicator employed and is small.

A TRULY portable and sensitive electrometer might be very useful to industry as well as to the laboratory physicist if it could be made sufficiently rugged to withstand the severe treatment to which most commercial instruments are subject. The use of ordinary suspension type laboratory instruments on shipboard is quite impossible and in certain problems it has been necessary to develop a sensitive electrostatic instrument which is sufficiently rugged to use either on shipboard or on aircraft. In this paper we give the underlying principles of the device which are believed to be new and describe one of the early models of a practical instrument.

The instrument we shall describe has been developed to replace the widely advertised vacuum tube type of electrometer, because of certain difficulties encountered in its general application. The vacuum tube device, primarily because of its simplicity, is quite satisfactory in an isolated laboratory where shielding of the entire system, galvanometer and special batteries is practical but the tube is especially susceptible to radio interference and in ordinary service many difficulties arise. Moreover, its voltage sensitivity is not great when used with ordinary portable indicating meters and its leakage resistance is not a constant.

With the exception of the vacuum tube electrometer, all the usual electrostatic devices employ light suspended mechanical systems which are acted on by electrostatic charges. Evidently these light suspended systems make such instruments commercially impractical and while the new device is not yet superior in sensitivity to a sensitive quadrant electrometer like the Compton instrument, it can undoubtedly be made so, and has the added advantage of complete portability. The new instrument is a true electrometer and the charges placed on the semicylinders are free of all leakage effects, save only the usual small leakage through the amber supports and through the surrounding air. By evacuating the cylindrical system the latter leakage may be avoided.

Principle

The underlying principle of the new instrument may be understood from a consideration of the simplified circuit of Fig. 1. In this circuit the capacity C_1 represents a fixed capacity of the system upon which is placed an unknown charge q_0. The capacity of the condenser C_2 is made periodic and takes on all values from zero to C_2'. Initially when the charge q_0 is on C_1, C_2 is zero but as C_2 is increased the condenser C_1 shares its charge with C_2 and a charge flows through the wires that connect the condensers and hence through the inserted resistance R. The current through the resistance sets upon a potential difference across R which can be measured or amplified by a vacuum tube, $V.T.$ As soon as the capacity of the condenser C_2 starts to decrease in value, the current through R is reversed and the charge flows back into C_1. The system constantly strives to reach the static state in which the charges are divided in proportion to the capacity but this condition is never attained because C_2 constantly changes with time.

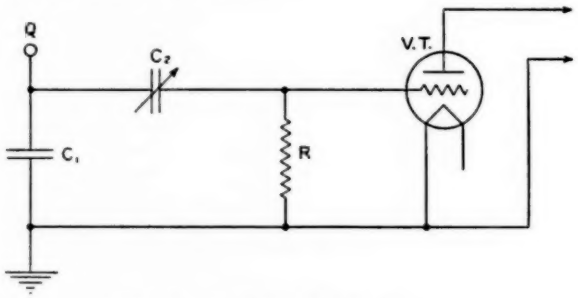

Fig. 1. Simplified circuit diagram.

It is to be especially noted that after C_2 had returned to its original zero capacity, the entire initial charge q_0 is back on the condenser C_1 and the interchange of charge between the two condensers has added or subtracted nothing to the electrical energy because the energy dissipated in the resistance R was supplied by the mechanical forces which make C_2 periodic. The net result of the interchange of charge between the two condensers is an alternating potential across R of a frequency equal to that of the variation of C_2 and this potential is proportional to the original static charge. Thus the circuit may be adopted for measurement of static charge by simply indicating the magnitude of the alternating potential difference. Moreover, since the generated potential is alternating it may readily be amplified a great many times and the sensitivity can undoubtedly be pushed up to a point over a hundred times that already attained.

If at the time $t=0$, we take $C_2=0$ and $q_1=q_0$ then the circuital relations lead to the integral equation

$$\frac{q_0}{C_1} - Ri - \frac{\frac{C_1}{C_2}+1}{C_1}\int i\,dt = 0 \qquad (1)$$

where i is the current through the resistance R and q_0 is the initial charge on C_1. This equation is not readily solved except by approximation because C_2 changes with time. We choose to solve Eq. 1 for the potential E applied to the grid of the vacuum tube and leave it in the integral form. Thus

$$E = \frac{q_0}{C_2}\left[\frac{q_1}{q_0}\left(1 + \frac{C_2}{C_1}\right) - 1\right] \tag{2}$$

where q_1 is the instantaneous charge on C_1. In the steady state it is evident from Eq. 2 that E will be zero but as C_2 changes periodically as the result of mechanical forces E will become finite and periodic with C_2. We can see without analysis that the potential E across R will increase as dq/dt increases and therefore at low frequencies will be approximately proportional to the frequency of the variation of C_2. However, when the frequency approaches the relaxation frequency of the circuit the condenser systems have insufficient time to completely charge or discharge and E drops. In general, the maximum E for a given applied charge q_0 occurs when the frequency of the change in C_2 is slightly less than f_r the effective relaxation frequency of the circuit. Thus

$$f_r = 1 + \frac{C_1}{\bar{C}_2} \Big/ RC_1 \tag{3}$$

where $\bar{C}_2$ is a mean value of the periodically variable capacity of the condenser C_2. The requirement in regard to the driving frequency need only be met when maximum sensitivity is desired.

Method

The potential applied to the grid of the first vacuum tube is undoubtedly of very complicated wave form but it will necessarily have a strong component of the fundamental frequency at which C_2 is varied. This fundamental alternating potential can be amplified and, by rectification at the output by an electron tube or by a copper oxide rectifier, can be made to operate a sensitive portable direct current instrument.

Such a method was used initially but it proved to be entirely unsatisfactory because: (1) amplifier noises are rectified equally with the impressed signal and the indicating meter measures these; (2) the arrangement is incapable of distinguishing the sign of the charge q_0. In the special problem for which the instrument was designed it was necessary to know both the magnitude and the sign of the unknown change, and a new system has been devised to attain this goal.

It is well known that vacuum tube amplifiers are poorly adapted for the amplification of steady potentials so we have avoided their use in this manner by a special system which is applicable to a great many electrical problems. Starting with a steady potential we convert this into an alternating potential, as in the case described above, by a special inverter unit. The resulting alternating potential is amplified to any desired degree, by suitable arrays of vacuumtubes, and instead of measuring the resulting a. c. potential directly,

it is passed through a converter operating in strict synchronism with the inverter unit. This rectifies the output a.c. and it is led directly to a direct current indicator. It is evident that a reversal in sign of the applied potential will reverse the phase of the amplified a.c. and the direct current indicator will reverse its sign. Thus the arrangement is capable of not only measuring the magnitude of the applied potential but also its sign. The system has other valuable properties; for example, fluctuations in the amplifier or a moderately high noise level due to microphonic effects will not appreciably affect the indicator because the indicator obviously responds only to the inverter-converter frequency. The amplification per stage can therefore be carried to quite high values. Synchronous mechanical rectifiers are highly efficient and, aside from their occasional erratic operation due to the poor condition of the contacts, are quite satisfactory. It is to be noted that the indicating meter and the control mechanism may be readily placed at any desired distance from the electrometer proper which permits the device to be installed in inaccessible positions. In certain applications this is of considerable importance.

An Instrument

The general design of an early model is shown in Fig. 2, in which 1 represents the driving motor of sufficient power to drive the inverter-converter unit at a constant speed. Semicylindrical electrodes 4 and 5 take the place of

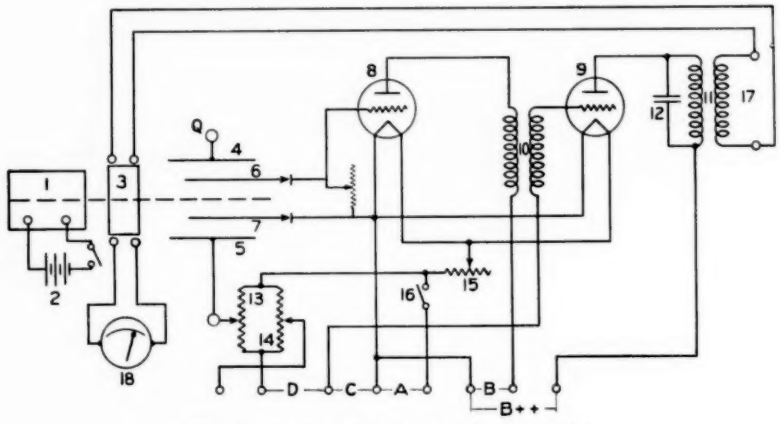

Fig. 2. General design of an early model.

the usual quadrants and one of them (4) is highly insulated by means of amber. The semi-cylindrical inductors 6 and 7 revolve inside the electrodes and are suitably spaced from them. The inductors are connected to the input of a vacuum tube amplifier 8, 9 and 10 and the alternating current induced in them, due to the presence of static charges on the electrodes, is amplified and passed through a converter unit 3 by way of an output transformer 11.

The inverter and converter units are driven by the same shaft and therefore the current delivered to the indicator 18 is always direct current. The converter unit is simply a two segment copper commutator, each segment

being connected to a ring. The brush assembly is capable of being shifted about the axis of rotation to permit the proper adjustment of phase. A potentiometer 13 is provided to adjust properly the potential of one semicylindrical electrode and a magnetic switch is used to bring the electrode 4 to the potential of 5 during periods of adjustment. Suitable batteries are connected to the system as indicated in the diagram.

Except under unfavorable conditions it has been only necessary to shield and ground the inverter unit together with the lead to the first grid. All battery and meter leads of ordinary twisted conductor have been run wherever necessary without shielding.

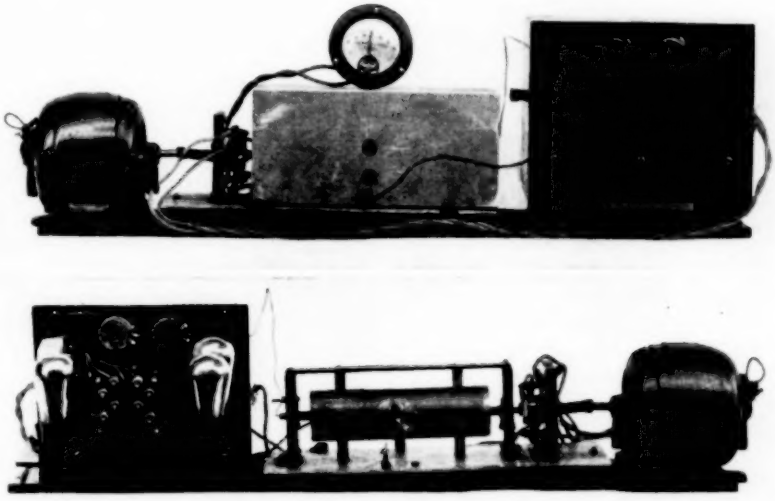

Fig. 3. Photographs of complete instrument.

Sensitivity

The sensitivity of the electrometer may be varied over any desired range by simply changing the magnitude of the resistance between the inductors 6 and 7. The instrument shown can be operated at any sensitivity up to about 10,000 microamperes per volt. A slight fluctuation of the indicating meter, due to commutation, amounting to about one microampere limits the effective voltage sensitivity with portable instruments to about 10^{-4} volts. Higher sensitivity can undoubtedly be secured by paying closer attention to details of construction and increasing the overall amplification.

The capacity of the electrometer can be controlled over a moderate range. The one described above had a capacity of 44 centimeters and therefore its useful quantity sensitivity with portable instruments was 4×10^{-15} coulomb per microampere. Experience with the device indicates that instruments with capacities as low as 10 cm could be built without sacrificing the voltage sensitivity indicated above.

The period of the instrument depends on the meter used as an indicator and is small.

Conclusion

The difficulties encountered in the development of the instrument have been largely those of mechanical design. Present instruments are satisfactory and retain their calibration over considerable periods of time. It is necessary to clean occasionally the converter unit and check the over-all gain of the amplifier to insure the stability of the zero. We have not attempted to push the sensitivity of the instrument to its logical limit, about 100 times greater, because there has been no need for such high sensitivities. The present device is comparable in sensitivity to a Compton electrometer in stable adjustment, yet is completely portable and sufficiently rugged to be used in aircraft.

LETTERS TO THE EDITOR

Prompt publication of brief reports of important discoveries in physics may be secured by addressing them to this department. Closing dates for this department are, for the first issue of the month, the twenty-eighth of the preceding month; for the second issue, the thirteenth of the month. The Board of Editors does not hold itself responsible for the opinions expressed by the correspondents.

Rotational Structure and Predissociation of the P_2 Molecule

In a recent issue of this journal Jenkins and Ashley[1] in a short abstract have come to the conclusion that the nuclear spin of the P^{31} nucleus is $\frac{1}{2}$, the intensity ratio of successive lines in a P_2 band being 3:1. Some time ago I published a short note in Nature,[2] in which I showed that some of the P_2 emission bands suddenly break off at a certain low value of J in consequence of predissociation. The plates were taken with a two-meter grating at Bristol which gave extremely good definition (line width on the plate <0.01 mm) so that the fine structure of the bands was quite well resolved. But in consequence of ghosts of the strong lines and overlapping of other bands the question whether the bands showed intensity alternation or not could not be decided by mere inspection of the plates. It was only possible to say that there was either a strong alternation (1:3) or none. The exact measurement and analysis of the plates was unfortunately delayed rather long largely because the comparator necessary for it has been supplied only very recently. The measurement of the clearly resolved part of the spectrum with the new Abbe comparator was just finished when I saw Jenkins and Ashley's abstract. Though these authors state that they are going to undertake a detailed analysis of the rotational structure I nevertheless decided to analyse that part of the spectrum I had measured in view of the rather large amount of work put into this matter and because I wished further to investigate the question of predissociation of P_2 which I had initiated.

The result of my rotational analysis is in agreement with Jenkins and Ashley's conclusion that there is an intensity alternation 3:1. The weak lines are only partly to be seen on my plates. I have measured only the strong ones. As already stated by Jenkins and Ashley the bands have only one P and one R branch as is characteristic of a $^1\Sigma-{}^1\Sigma$ transition. The relative numbering of the lines in the bands is at once given by the breaking off of the branches in some of the bands. In the bands with $v'=11$ (according to Jakowlewa's[3] assignment of vibrational quantum numbers) the breaking off of the strong lines of the P and R branches occurs at $J'=34$, in those

[1] F. A. Jenkins and M. Ashley, Phys. Rev. **39**, 552 (1932).

[2] G. Herzberg, Nature **126**, 239 (1930).

[3] A. Jakowlewa Zeits. f. Physik. **69**, 548 (1931).

TABLE I.

		B	
$^1\Sigma_g^+$	$v''=0$	0.311 (extrapolated)	
	$v''=28$	0.258_0	$\alpha''=0.001_9$
	$v''=29$	0.256_1	$r_0''=1.86$A
	$v''=30$	0.254_1	
	$v''=31$	0.252_2	
$^1\Sigma_u^+$	$v'=0$	0.236_5 (extrapolated)	
	$v'=9$	0.224_2	$\alpha'=0.001_4$
	$v'=10$	0.222_8	$r_0'=2.14$A

with $v'=10$ at $J'=58$. The combination relations for the bands with equal v' or v'' fitted usually within less than 0.2 cm^{-1}. As only bands with high v' and v'' have been analysed the B_0 values cannot be given with high accuracy. In Table I are therefore given the B values of the bands actually analysed as well as B_0' and B_0'' which have been got by linear extrapolation. The former are accurate to 0.001 cm^{-1}. In the analysis it was assumed that the rotational constant D is negligibly small.

In Table II the B_0 and r_0 values for the ground states of the molecules N_2, O_2, P_2[4] and S_2 are given for comparison. It is very remarkable and most anomalous that the B_0 value of P_2 is much lower and the r_0 value correspondingly higher than that of S_2 so much the more as the vibrational frequency is slightly larger. As N_2 has a higher B_0 than O_2 one would of course expect the same for P_2 and S_2. Neither the empirical formula of Morse for B nor that of Birge for α hold in this case.

TABLE II.

	B_0''	r_0''
N_2	1.992	1.10
O_2	1.4375	1.21
P_2	0.311	1.86
S_2	0.408	1.60

As the P_2 band system under investigation evidently corresponds to the far ultraviolet band system of N_2 (the Lyman-Birge-Hopfield bands) the latter presumably also represent a $^1\Sigma-{}^1\Sigma$ transition and not as hitherto assumed $^1\Sigma-{}^1\Pi$. This is in better agreement with the electron configuration of N_2.

[4] That the lower level of the band system discussed here is the normal level follows from the fact that it occurs in absorption also.

[5] G. Herzberg, Ergebn. d. ex. Naturwiss. **10**, 207 (1931).

From the fact that the strong lines measured by me have odd J'' it follows that the P^{31} nuclei follow the Fermi statistics, if one makes the plasuible assumption that the ground state of P_2 like that of N_2 is $^1\Sigma_g^+$. There is however a remote possibility that the numbering of the lines is to be altered, if D is not negligible. This I could not decide until now.

The last observed strong rotational level ($J=58$) of the state $v'=10$ has an energy which is 73 cm^{-1} *higher* than the last observed strong level ($J=34$) of the state $v'=11$. This is in agreement with the considerations mentioned in my report on predissociation.[5] It follows from the fact that for higher values of J there is a higher maximum of the curves of effective potential energy to be overcome.

As shown by Jakowlewa the molecule predissociates into one normal (^4S) and one excited (^2D) P atom. As a ^4S and a ^2D atomic state cannot form a singlet molecular electronic state it follows that the predissociation of P_2 is a *radiationless intercombination*. This seems to be quite certain. As according to Kronig intercombinations are forbidden in predissociation just as in transitions with radiation the transition probability will be very low so that possibly in absorption no diffuseness of the corresponding bands will be observable. It is significant that though the rotational levels lie very close together in this heavy molecule the breaking off of the band occurs from one line to the next. This can only be explained if one assumes case Ib in the author's nomenclature (l.c.) i.e., the point of intersection of potential curves below the asymptote of the dissociated state.

A full paper on this subject will appear shortly.

G. HERZBERG

Physikalisches Institut der
Technischen Hochschule,
Darmstadt, Germany,
March 9, 1932.

The Efficiency of Electron Emission by Metastable Atoms

In connection with the recent paper of Langmuir and Found,[1] it seems important to draw attention to certain observations by Penning which make it possible to derive an upper limit to the probability λ that a metastable atom impacting on a metal surface will release an electron. In the case of neon metastables impacting on a degassed Fe surface with energy of thermal agitation only, $\lambda<0.005$, an order of magnitude less than the photoelectric efficiency of neon resonance radiation.[2]

[1] I. Langmuir and C. G. Found, Phys. Rev. **39**, 237 (1932).

[2] C. Kenty, Phys. Rev. **38**, 377(L) (1931).

In the course of his studies on the effects of small admixtures of other gases on the sparking potential in the noble gases,[3] Penning observed that the sparking potential in pure neon is changed by less than 5 volts when the spark gap is intensely illuminated by the light from a neon positive column discharge.[4a] That this illumination actually reduces the number of metastable atoms present very greatly (to about 1/5, from a comparison of Fig. 4[4] with Fig. 6[5]) is shown by the increase of the sparking potential in Ne+0.001% A when the gap is so illuminated.[4] The pressure variation of the S.P. shows that this irradiation does not produce any secondary products which release electrons from the cathode.

Similarly, it can be shown that the addition of N_2 to A removes metastable A atoms without producing ions,[6] but the addition of N_2 to pure A actually reduces the S.P. slightly.[6a] Apparently λ is as small for A as for Ne, although the data necessary to a quantitative computation are lacking.

If Penning's conditions are approximated by assuming the plane electrodes to be parallel and infinite, the condition for equilibrium of metastables in the discharge is

$$bN - D(d^2N/dx^2) = f\alpha e^{\alpha x}, \tag{1}$$

where N is the metastable concentration (reduced to a current density of 1 electron per cm² per sec.), D is their diffusion coefficient, b is the inverse of the average life of a metastable atom, α is the number of ions formed by an electron in moving 1 cm in the direction of the lines of force (Townsend's α), f is the ratio of the number of metastable atoms formed to the number of ions formed, and x is distance from the cathode.

The complete solution of (1), under the condition that $N=0$ at $x=0$ and at $x=1$ (the plates in Penning's experiment were 1 cm apart), is

$$\begin{aligned} N = {} & ce^{\alpha x} + [c(e^{\alpha-u} - 1)e^{-ux}/(1 - e^{-2u})] \\ & - [ce^{-2u}(e^{\alpha+u} - 1)e^{ux}/(1 - e^{-2u})] \end{aligned} \tag{2}$$

where $u=(b/D)^{1/2}$, and $c=\alpha f/(b-D\alpha^2)$. The number of metastables striking the cathode is

$$\begin{aligned} n &= D(dN/dx)_{x=0} \\ &= D[\alpha c - \{uce^{-2u}(e^{\alpha+u} - 1)/(1 - e^{-2u})\} \\ &\quad - \{uc(e^{\alpha-u} - 1)/(1 - e^{-2u})\}] \end{aligned} \tag{3}$$

Substituting $D=44$,[7] $b=1290$,[8] $f=2.7$,[9] $\alpha=4.6$,[10,4a] we get

$$n = 8.3 \tag{4}$$

The condition for equilibrium of electrons in the discharge is

$$\gamma(e^{\alpha} - 1) = 1, \tag{5}$$

where γ is the probability that an ion impacting on the cathode will release an electron. Differentiating with respect to V, the S.P.

$$d\gamma = -e^{\alpha}(d\alpha/dV)dV/(e^{\alpha} - 1)^2. \tag{6}$$

The number of positive ions striking the cathode is

$$n' = (e^{\alpha} - 1). \tag{7}$$

If n metastables also strike the cathode, they will release $n\lambda$ electrons, and due to them

$$\begin{aligned} d\gamma &= n\lambda/n'; \\ \lambda &= -e^{\alpha}(d\alpha/dV)dV/(e^{\alpha} - 1). \end{aligned} \tag{8}$$

Substituting $(d\alpha/dV)=0.019$[10] and $-dV<5$,[4a]

$$\lambda < 0.0023.$$

Allowing for the various uncertainties introduced, we may write with certainty

$$\lambda < 0.005.$$

In consequence, the results of Langmuir and Found,[1] Uyterhoeven and Harrington,[11] and others, must be interpreted in some other manner, perhaps photoelectrically, as proposed by Kenty.[2]

EUGENE W. PIKE

At present,
Seminar f. theor. Physik,
Universität Greifswald.
March 16, 1932.

[3] F. M. Penning, (Review) Phil. Mag. **11**, 961 (1931).
[4] F. M. Penning, Zeits. f. Physik **57**, 723 (1929).
[4a] F. M. Penning, ibid, p. 729.
[5] F. M. Penning, Zeits. f. Physik **46**, 335 (1928).
[6] F. M. Penning, Physica **10**, 47 (1930).
[6a] F. M. Penning, ibid, p. 56.
[7] M. W. Zemansky, Phys. Rev. **34**, 223 (1929).
[8] F. M. Penning, reference (5), p. 342.
[9] F. M. Penning and M. C. Teves, Physica **9**, 97 (1929).
[10] J. S. Townsend and S. J. McCollum, Proc. Roy. Soc. **A124**, 533 (1929).
[11] W. Uyterhoeven and M. C. Harrington, Phys. Rev. **36**, 709 (1930).

A Plausible Explanation of the Alkali Inverted Doublets

It is well known that the $^2D_{3/2,5/2}$ and $^2F_{5/2,7/2}$ terms of the heavier alkali atoms are inverted. For these doublets the term with the largest J-value lies deepest. In view of the well-known perturbation-resonance phenomenon that occurs between similar energy levels, treated quantum mechanically by Heisengberg, Darwin, Pauli, Slater, Condon, Langer, and others, and of the recent paper by Shenstone and Russell on "Perturbed Series in Line Spectra" (Phys. Rev. **39,** 415, 1932) a plausible explanation of the inverted doublets suggests itself. The $^2D_{3/2,5/2}$ (*even*) and the $^2F_{5/2,7/2}$ (*odd*) terms in the alkalies should not only be normal, i.e., smallest J-value deepest on an energy level diagram, but very narrow as well. If in the neighborhood of the $^2D_{3/2,5/2}$ (*even*) terms other $^2D_{3/2,5/2}$ (*even*) terms occur and these latter terms are inverted there will be an interaction between the two sets of levels, (i.e., between levels having the same J-values) tending (1) to invert the normal doublet, and (2) to right the inverted doublet. In general this tendency to invert the normal doublet will decrease with the distance between the two sets of levels and will increase with the doublet separation of the perturbing levels. If this perturbing inverted doublet has a large separation the effect will be to invert the normal $^2D_{3/2,5/2}$ terms. Such a wide inverted ^2D term occurs in potassium for example by the excitation of a $3p$ electron from the completed $3p^6$ subshell to a $4p$ state. Even though the resultant electron configuration $3p^54s4p$ should lie somewhere between 50,000 to 100,000 cm^{-1} above the perturbed levels, the wide separation of the inverted ^2D terms makes the inversion of the perturbed levels seem probable. By making certain justifiable approximations the magnitude of the interactions can be calculated on the quantum mechanics. This work is now being carried on in this laboratory by Miss M. Phillips.

H. E. WHITE

Department of Physics,
University of California, Berkeley,
March 19, 1932.

The Ionization Potential of Radon

Because of the rarity of the inert gas radon formed from the radioactive disintegration of radium, it was thought that the experimental determination of the ionization potential of the element would be impossible. Hence, estimates of the value of this constant were made by extrapolating different functions correlating the spectral properties of the noble gases. Turner[1] suggested a value of 27.5 ± 1.5 volts based on a linear relationship which he had observed between the atomic number of the elements and their quantum defects. By applying Eve's rule[2] that the product of the ionization potential V and the ionic radius r is roughly a constant for the members of a periodic sub-group, Glocker[3] deduced that the ionization potential of radon is 10.0 ± 0.9 volts by evaluating the radius of the atom (2.80 ± 0.25A) from the measurements of the diffusion coefficient of the gas in air. In a critique on Glocker's assumptions, Struwe[4] concluded that a probable value of the ionization potential is 9 volts. Independently, Piccardi[5] calculated a value which was equal to 11.96 volts on the basis of a generalization which he had observed by plotting the difference between the ionization potentials of the elements of adjacent groups having the same total quantum number n against their atomic numbers. Biswas[6] suggested that the ionization potential of radon will exceed that of xenon, as the last member of practically all the other periodic groups exhibits a similar behavior with respect to the progression of the ionization potentials of the atoms within the group. By assuming an inverse proportionality between the ionization potential and the square of the mean atomic radius, he secured a value equal to 14.0 ± 0.5 volts depending upon the magnitude of the radius which he estimated as 1.8 ± 0.03A from the known viscosity and critical constants of the gas.

The spectrum of radon has recently been

[1] L. A. Turner, Phil. Mag. **48,** 384, 1010 (1924).

[2] A. S. Eve, Nature **107,** 552 (1921); Phys. Rev. **27,** 515 (1926).

[3] G. Glocker, Phil. Mag. **50,** 997 (1925).

[4] F. Struwe, Zeits. f. Physik **37,** 859 (1926).

[5] G. Piccardi, Atti. accad. Lincei (6) **6,** 305, 428, 621 (1927).

[6] S. C. Biswas, Phil. Mag. (7) **5,** 1094 (1928).

analyzed into its terms by Rasmussen.[7] His analysis reveals that the ionization potential of the neutral radon atom is 10.689 volts. Shortly afterwards Holweck and Wertenstein[8] measured the ionization potential of the gas by applying the well-known method of Hertz for the study of excitation potentials to 300 millicuries of radon in a tube where its pressure amounted to about 0.8 bar. Their value of 10.6 volts establishes the true magnitude of this constant, as experiments with equally small quantities of krypton and xenon also yielded ionization potentials that were in good agreement with the accepted values.

In a communication to the Physical Review the present writer[9] has evaluated the energy levels of the 1^2S and m^2P_j terms of the neutral atom of atomic number 87 and those of the singly ionized radium atom with the aid of a rule stating that within a periodic sub-group the quantum defect of a spectral level is approximately a linear function of the total quantum number of the valence electron. As the ionization potential can be calculated directly from the normal term, the quantum defect rule is also applicable to the coordination of the ionization potentials of the elements. In Table I the calculated quantum defects and ionization potentials of the noble gases are compared with the observed values. The quantum defects were estimated from the relationship: (1) $Q=an+b$, in which the coefficients a and b were evaluated by the method of average points on the data for only neon, argon, krypton and xenon. The calculated values of helium and radon therefore represent extreme extrapolates of Eq. (1), and the fair agreement with the observed constants proves the general validity of the equation.

In view of the researches of Rasmussen and of Holweck and Wertenstein there is little doubt but that the ionization potential of the neutral radon atom is 10.7 volts. The fact that this value follows the normal trend of the ionization potentials of the noble gases is surprising, since most of the spectroscopic properties of the atoms occurring in the seventh period exhibit anomalous behavior in this respect. Thus in the case of the singly ionized radium atom, the doublet separations, the wave-numbers of the principal series spectrum and the term values as arranged by Hicks[10] have properties which are considerably different from those expected from a simple extrapolation of the corresponding values for the other alkaline earth atoms. Indeed, to account for the 1^2S, 2^2P_2 and 2^2P_1 levels of Ra II it was found necessary to introduce a corrective term ϵ in the formula employed for their evaluation.[9] A more complete analysis of the arc and spark spectra of the radium atom than is at present available is desirable for the better understanding of the unusual behavior of seventh-period atoms.

TABLE I. *Quantum defects and ionization potentials of the noble gases.*

$$Q = 0.9196n - 0.6607 \qquad (1) \qquad V = \frac{300hcR}{e(n-Q)^2} = \frac{13.538}{(n-Q)^2} \qquad (2)$$

Element	n^*	n	Quantum defect calc. (1)	Quantum defect obs.	Ionization potential calc. (2)	Ionization potential obs.
Helium	0.7438	1	0.259	0.256	24.7	24.47[a]
Neon	0.7941	2	1.179	1.206	20.1	21.47[b]
Argon	0.9292	3	2.098	2.071	16.6	15.68[b]
Krypton	0.9858	4	3.018	3.014	14.0	13.940[c]
Xexon	1.0591	5	3.937	3.941	12.0	12.078[c]
Radon	1.1254	6	4.857	4.875	10.4	10.689[d]

[a] T. Lyman, Nature **110**, 278 (1922).
[b] Landolt-Börnstein Physikalisch-Chemische Tapellen, 5th ed., vol. 2, table 152f, Berlin (1931).
[c] Meggers, de Bruin and Humphreys, Bur. Stand. J. Research **3**, 129, 731 (1929).
[d] E. Rasmussen, Zeits. f. Physik **62**, 494 (1930).

[7] E. Rasmussen, Zeits. f. Physik **62,** 494 (1930).

[8] F. Holweck and L. Wertenstein, Nature **126,** 433 (1930).

[9] H. Yagoda, Phys. Rev. **38,** 2298 (1931).

[10] W. N. Hicks "Analysis of Spectra", pg. 276 Cambridge University Press 1922. A. Fowler "Report on Series in Line Spectra", p. 138, London 1922.

HERMAN YAGODA

Sea Gate, N. Y.
March 22, 1932.

Positive Ions of Mass 220

It is a well-established fact that when alkali sulfates are heated, they give off singly charged atoms of the alkali metals. The sulfates provide good sources of positive ions of the alkalies and do not volatilize too rapidly when heated to a sufficiently high temperature to give a good ion emission. With this in mind the authors undertook a mass-spectrograph examination of a mixture of sulfates of the alkalies which is believed to contain a small amount of the sulfate of element 87. The sample, which was provided by Professor Jacob Papish and Mr. Eugene Wainer, was derived from the alum used by them in obtaining x-ray lines due to element 87. The alum is described in their paper which appeared in the Journal of the American Chemical Society **53,** 3818 (1931).

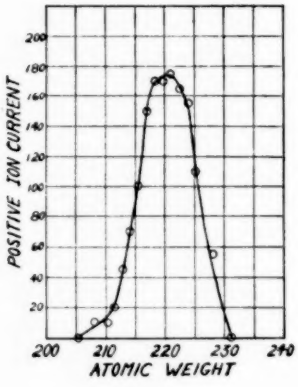

Fig. 1.

The sample was heated on platinum in a "Dempster Type" mass spectrograph at a pressure less than 10^{-6} cm Hg. At a temperature between 600 and 700°C a small but fairly steady positive ion emission was obtained in the region of the mass scale corresponding to singly charged atoms of atomic weight 220. Fig. 1 shows one of six curves which were obtained under good working conditions, the mass scale being checked by the caesium peak from the same source at the same temperature. The positive ion currents shown here are in arbitrary units but are of the order of 10^{-15} to 10^{-14} amperes. The caesium ion currents from the same source at the same temperature are about 10,000 times larger.

The average m/e value obtained from the six curves is 220, the maximum variation from this value being no greater than 1 part in 220. Thus the best value of m/e for these ions which can be given at the present stage of the work is 220 ± 1.

Further studies of these ions are in progress, and a more accurate determination of the m/e value will be undertaken in the near future.

The authors wish to thank Professor Papish and Mr. Wainer for their cooperation in this work.

LEROY L. BARNES AND R. C. GIBBS
Department of Physics,
Cornell University,
March 23, 1932.

Is γ-Radiation Produced by Proton Transitions in the Nucleus?

By using the formula

$$\tilde{\gamma}(\text{cm}^{-1}) = R'Z^2 \frac{P}{(t'+1)^2}\left(\frac{1}{t_1^2} - \frac{1}{t_2^2}\right)$$

where R' is a constant for all elements having a value of 2.017×10^5 cm^{-1}, P is the total number of protons present, t' is the total quantum number of the proton being excited; (See Bryden, Phys. Rev. **38,** 1989, 1931; and White, Phys. Rev. **38,** 2073, 1931), and t_1 and t_2 are quantum numbers similar to n_1 and n_2 in the atomic case, it has been possible to compute the values of $\tilde{\nu}$ and λ given in Tables I and II. The values of P were ob-

TABLE I.

$$\tilde{\gamma} = R'Z^2 \frac{P}{(t'+1)^2}\left(\frac{1}{1^2} - \frac{1}{2^2}\right) \qquad R = 2.017 \times 10^5 \text{ cm}^{-1}.$$

Element	Cal. ν	Cal. λ	Obs. λ	t'	Z
UX_1	7.22×10^5cm^{-1}	138.5 X.U.	134.2 X.U.	6	1
RaD	3.92	255	261	8	1
RdTh	7.04	142	145	6	1

tained from tables on page 24 of Rutherford, Chadwick and Ellis' book "Radiations from Radioactive Substances". The observed values of λ for the γ-radiation were taken from tables in the same book.

The observed λ's in Table I are the longest wave-lengths listed for the elements concerned, and are the strongest, or one of the strongest lines in the γ-ray spectrum of those elements.

The agreement between the computed and observed λ's is good and most surprising, perhaps, is the fact that the AcX lines can be represented by a spectroscopic series formula where t_1 takes on values of 3/4, 7/8, 15/16, 31/32, $\cdots$, 1. This is quite a different form from the series in optical spectra.

The intensities of the lines are peculiar.

Other radioactive spectra are being examined and a detailed report of all data will be published soon.

TABLE II. *γ-ray spark spectrum of AcX, AcX II.* $Z=2$; $t'=10$; $t_2=2$.

t_1	Cal. $\tilde{\nu}$	Cal. λ	Obs. λ	Int. obs.
3/4	$22.6\times10^5\text{cm}^{-1}$	44.2 X.U.	45.9 X.U.	130
7/8	15.63	64.0	61.7	25
15/16	13.1	76.3	78.6	80
31/32	12.0	83.3	80.6	180
1	11.10	90.1	86.0	155

The AcX lines in Table II include all the AcX lines listed by Rutherford, Chadwick and Ellis. The intensities listed are estimated intensities from the intensity of β-ray lines.

P. G. KRUGER
Department of Physics,
University of Illinois,
Urbana, Illinois,
March 24, 1932.

Probable Values of e, h, e/m and α,—An Addition

My complete article on this subject (to be denoted G. C. 1932) appears elsewhere in the present issue. The purpose of this letter is to make two corrections and also two additions.

(1) In discussing Bearden's grating measurements of x-rays, I used accidentally the values given by Bearden in a preliminary paper, footnote 30 of G. C. 1932. Fortunately the results given in his full paper[1] differ from the preliminary values by an entirely negligible amount.

(2) It was stated in G. C. 1932 that Bond, by using only my Eq. (1), failed to carry the new method to its logical conclusion as embodied in Eq. (2). It is true that Bond used only Eq. (1), but I overlooked the fact that, in a footnote, page 633 of reference 3, G. C. 1932, he remarked "By assuming a value of h, and plotting all the deduced values of e against $1/m$, the direct estimates of e could also be plotted at $1/m=0$. The increase in accuracy would, however, be negligible." This is essentially a statement of Eq. (2), which Professor Brode quite independently suggested to me, as noted in G. C. 1932.

(3) I remarked in G. C. 1932, that the true value of e/m might well be 1.760 or even lower but it seemed better to retain, for the present, the 1929 spectroscopic value of 1.761. Since this was written new evidence has appeared that favors such a lower value. Kirchner[2] corrects his previous determination, 1.7598 ± 0.0025, to 1.7585 ± 0.0012, and from new measurements obtains 1.7590 ± 0.0015. In order to show explicitly the dependence of the various constants on the value adopted for e/m, I have made a new solution (to be called solution *l*) that is similar to my final solution *k*, except that one assumes $e/m=1.759\pm0.001$ in place of 1.761 ± 0.001. The resulting probable errors are again based on internal consistency ($r_e/r_i=0.571$), and are therefore identical with the errors of solution *k*, since the weights of the data are unchanged. The results are $h=6.5442\pm0.0091$, $e=4.7677\pm0.0040$, $1/\alpha=137.369\pm0.048$, $e/m=1.7591\pm0.0009$.

From solutions *k* and *l* one can now obtain, by mere linear interpolation, values of the constants that correspond to any intermediate value of e/m. The value of h is thus seen to be, over this range, almost independent of the

[1] J. A. Bearden, Phys. Rev. **37**, 1210 (1931).
[2] F. Kirchner, Ann. d. Physik **12**, 503 (1932).

value of e/m, and the various possible values of e all lie very close to the oil-drop value 4.768 ± 0.005. It may be of interest to note that the value $1/\alpha = 137.348$, given by the theory of ultimate rational units[3] corresponds to $e/m = 1.7597$.

(4) Dr. Bond has kindly sent me the advance proof of a paper that he presented to the London Physical Society on February 19, 1932. In this paper Bond discusses, among other things, Eddington's most recent theory[4] concerning the ratio of the mass of the proton to that of the electron ($M_P/m = R$). This theory, which was not mentioned in G.C. 1932, gives R as the ratio of the two roots of the equation $10x^2 - 136x + 1 = 0$. The coefficients 10 and 136 denote the number of degrees of freedom of certain elementary systems. The resulting value of R is 1847.599464. From this value one can calculate e/m, since $e/m = (R+1)F/H^1$, where F is the Faraday, 9648.9 ± 0.7 abs. e.m. units, and H^1 is the atomic weight of the chief isotope of hydrogen. The value of H^1 is best calculated from Aston's measurement of the atomic mass, reduced to the chemical scale. Recent evidence from band spectra indicates that Aston's probable error can be taken conservatively as one-third his stated limit of error. The value of H^1 is thus 1.00756 ± 0.00005, and it was from this result that Birge and Menzel[5] predicted the existence of an isotope of hydrogen, of mass 2 and relative abundance $H^1/H^2 = 4500$. This isotope has subsequently been found,[6] with almost exactly the predicted abundance. The resulting value of e/m is 1.77031 ± 0.00014.

This value of e/m, used in connection with Bohr's formula for the Rydberg constant, gives one theoretical relation between e and h. Eddington's theory that $1/\alpha = 137$ yields a second relation between e and h. From these two theories one thus obtains $h = 6.5490 \pm 0.0011$, $e = 4.775855 \pm 0.000048$. The stated probable errors arise from the probable errors in F, H^1 and c. The error in R_∞ is negligible. In the Physical Society paper Bond calculates similar results, but he apparently did not consider the mass 2 isotope of hydrogen, so that his numerical results differ somewhat from mine. His probable error for $e(\pm 0.0004)$ is also much larger.

With these predicted values of e, h and e/m, one obtains $\sigma = 5.7365$, very close to my adopted value 5.735 ± 0.011, $c_2 = 1.4304$, fairly close to my adopted 1.432 ± 0.003, and $h_{3/3} = 6.5410$, close to 6.543, the best photoelectric result. Eddington's theories are, however, in definite disagreement with the newer experimental values of e/m, and in mild disagreement with the oil-drop value of e. The foregoing calculations have been made in order to show precisely the numerical consequences of these new theories, which Bond claims are in satisfactory agreement with existing data. The writer cannot, at this time, subscribe to such a conclusion.

RAYMOND T. BIRGE

University of California,
March 25, 1932.

[3] G. N. Lewis and E. Q. Adams, Phys. Rev. (2) **3**, 92 (1914).

[4] A. S. Eddington, Proc. Roy. Soc. **A134**, 524 (1931).

[5] R. T. Birge and D. H. Menzel, Phys. Rev. **37**, 1669 (1931).

[6] H. C. Urey, F. G. Brickwedde and G. M. Murphy, Phys. Rev. **39**, 164 (1932).

BOOK REVIEWS

Matière et Atomes. Second edition. A. BERTHOUD. Pp. 324, fig. 28. Gaston Doin and Cie, Paris. Price, 26 fr.

The claim of this little book on the attention of readers of the Physical Review is principally based on its pedagogical and educational significance. It forms one of the series *Encyclopédie Scientifique* issued for the consumption of the general public in France. It would be well if in America we could have similar works at such a modest cost without the usual expensive binding and "fixings" but with good, well-written, and reasonably up-to-date texts.

The ground covered is the standard material of atomic physics—x-rays, isotopes, radioactivity, Bohr theory and the periodic system, and an introduction to wave mechanics. The treatment is in the main descriptive but is adequate and to the point. This is one of the best elementary introductions to modern physics which the reviewer has seen.

E. L. HILL
University of Minnesota

Fortschritte und Probleme der Atom forschung. J. STARK. Pp. viii+112, 11 figs. Verlag von Johann Ambrosius Barth, Leipzig 1931. Price RM 4.80.

Professor Stark has collected in this volume three lectures and three reprints of former articles. These six items make up the six chapters of the book as follows: I. The important problems of present-day physics. II. Contributions to optics as related to atomic structure. III. New phenomena concerning atomic structure and light-emissions. IV. The "axial" nature of the valency-field of carbon and nitrogen atoms. V. Causality and the behavior of the electron. VI. The dogmatism of the modern theories of physics. These lectures constitute on the one hand a refutation of wave-mechanics, which Professor Stark does not like at all and on the other hand they constitute an attempt to present his own ideas of atomic structure, the process of light-emission and of chemical valency. The mathematic theory of the wave-nature of matter is completely wrong in the author's opinion and his adverse remarks are directed mainly towards Professor Sommerfeld who appears as the principal exponent of Schroedinger wave-mechanics. However it appears to the reviewer that Professor Stark has little to offer in place of the present physical theory of the nature of matter. He completely ignores the quantitative results that have been obtained. He offers only qualitative notions and arbitrary pictures of "axial" atoms. One wonders if the author will not change his views when he becomes acquainted with the work of Heitler and London, Slater and Pauling on the nature of chemical valency as portrayed by quantum mechanics. (See following review).

GEORGE GLOCKLER
University of Minnesota

Atomstrukturelle Grundlagen der Stickstoffchemie. J. STARK. Pp. v+44, 93 figs. Verlag von Johann Ambrosius Barth, Leipzig (1931). Price RM 3.90.

This monograph consists of two parts: I. General concepts of atomic structure and atomic binding and II. Modes of combination of the valency-electrons of the nitrogen atom. The author presents his ideas of the nature of the atom, for which he accepts the Rutherford model but modifies it in the sense that he considers the surface of the atom to be so constituted that only certain positions of the valency electrons on the surface are possible. To maintain the electron in equilibrium he postulates an "inner kinetic" force besides the electric force acting on the electrons. While he does not mind to postulate new forces whenever necessary he refutes all modern physical theories, especially wave-mechanics, and offers, however, only qualitative notions in their place. The text is profusely illustrated with detailed pictures of chemical compounds such as benzene and especially nitrogen compounds. Professor Stark accepts the chemist's notion of directed valency and his views are no advance in thought beyond the well

known theory of octet chemistry as discussed by Professor G. N. Lewis. It is hoped that Professor Stark can render a service to physical and chemical theory by his criticism for he may force the proponents of wave-mechanics to consider their own ideas still more carefully. (See preceding review.)

GEORGE GLOCKLER
University of Minnesota

General Physics for Home Economics Students. FRED G. PERSON. Pp. xx+298. John Wiley and Sons, New York, 1931. Price $2.50.

This text aims to give an abbreviated, college course in physics suitable for students of home economics. Part I (146 pages) treats the mechanics of solids and liquids and contains a short chapter on "Water for the Home" and another on "Sewage and its Disposal". The treatments of mechanics differs from that in an ordinary high-school text mainly perhaps in that it describes and illustrates somewhat more completely a few devices used in the home. Part II (123 pages) treats heat, electricity, sound, and light in a very incomplete and elementary manner.

The figures and drawings, with but few exceptions, are good. Nearly every chapter is followed by a set of good problems. Many discussions and descriptions are given in terms unintelligible to the student, and the text is marred by too many misstatements such as "A charge of electricity stationary or moving is always accompanied by a magneic field, . . . "; "An ordinary 25-watt tungsten lamp burns 25 watts per second, but in computing power consumption the lamp is allowed to burn one hour and the measure is called 25 watt-hours". This text cannot be considered as meeting the standard of any kind of a college course in physics.

ANTHONY ZELENY
University of Minnesota

Molecular Rays. RONALD G. J. FRASER. Cambridge Series of Physical Chemistry. Pp. 204+x, 78 figs. The Macmillan Company, New York, 1931. Price $3.75.

This monograph discusses the technique of production and measurement of molecular rays and their application as a tool in various fields of investigation. Written from the experimental standpoint it presents in some detail experimental results obtained by the use of molecular beams but makes no attempt to discuss their theoretical background, except insofar as is necessary for an understanding of the experimental technique. The book will be of interest to those who are engaged in research involving the use of molecular rays, as they will find in it many details of method and technique previously available only in brief papers, and to the scientist specializing in other fields who is interested in a summary of results obtained by this method and an indication of its possibilities. The author has himself contributed experimental results of some importance and has spent some time in Professor Stern's laboratory. His firsthand acquaintance with the problems which he discusses is evident. Occasionally the book suffers by the omission of significant details—but on the whole it is excellent both in its exhibition of the possibilities of the method and its discussion of technique. There is an interesting foreword by Professor Stern.

A. ELLETT
University of Iowa

The Theory of Spherical and Ellipsoidal Harmonics. E. W. HOBSON, Pp. xii+500. Cambridge University Press. The Macmillan Co. New York 1931. Price $11.50.

This is probably the most complete treatise on the subject which has yet appeared. The only volumes of at all comparable scope were written by Heine at much earlier dates. What particularly distinguishes Professor Hobson's monograph from the usual, briefer books is the great care and thoroughness with which spherical harmonics of unrestricted (i.e., non-integral) degree and unrestricted argument are studied. In fact, almost half the volume is devoted to these harmonics of very general type, which are, of course, not needed for the commoner physical problems, but which are by no means entirely devoid of applications. The opening four chapters are devoted to the more usual theory of the harmonics of integral degree, and

present this material in a very lucid and readable fashion. A tremendous number of formulas obeyed by the harmonics, and many alternative ways of representing them are included. In a volume as complete as this, it would have been very appropriate to include the formulas expressing the transformation of associated Legendre functions under a rotation of axes. Instead only the more restricted and usual transformation relations for Legendre polynomials are given, called the addition theorems by Professor Hobson. The more general formulas have been given by Wigner, Kramers and Ittmann, and others, and their coefficients are of particular mathematical interest as the "Darstellungskoefficienten" of the rotation group. These coefficients also prove to be the same as the wave functions of the symmetrical top and are intimately related to the Lamé functions for surfaces of revolution. The last chapter is upon ellipsoidal harmonics. It would here perhaps have been mentioned that the researches of Wang on the asymmetrical top have shown that the four rather unsymmetrical determinants for determining the Lamé functions can be very compactly and symmetrically expressed in terms of a single determinant which permits a fourfold factorization.

To summarize, the volume may be recommended as an unusually complete, and lucidly written treatise on the properties of harmonics. References to the literature and good indices are included.

J. H. Van Vleck
University of Wisconsin

CURRENT LITERATURE OF PHYSICS

Extracts from the Tables of Contents of Some of our Contemporaries

PROCEEDINGS OF THE ROYAL SOCIETY

VOL. 135, NO. A827, SERIES A. MARCH 1, 1932

Half-Value Period of Uranium-Y. By **O. Gratias and C. H. Collie** 299
Thermal Decomposition of Gaseous Diethyl Ether at High Pressures. By **D. M. Newitt and M. A. Vernon** 307
Thermodynamics of the Surfaces of Solutions. By **J. A. U. Butler** 348
Investigations in the Infrared Region of the Spectrum. Part V. The Absorption Spectrum of Carbonyl Sulphide. By **C. R. Bailey and A. B. D. Cassie** 375
Behavior of a Single Crystal of Aluminium under Alternating Torsional Stresses while Immersed in a Slow Stream of Tap Water. By **H. J. Gough and D. G. Sopwith** 392
Polarization of Electrons by Double Scattering. By **N. F. Mott** 429
Eigenfunctions for Calculating Electronic Vibrational Intensities. By **P. M. Davidson** 459

PHILOSOPHICAL MAGAZINE

VOL. 13, NO. 85 MARCH, 1932

An X-ray Investigation of the Crystals of o-Azataluene. By **Mata Prasad and K. V. Desai** 600
Equation of Motion of a Viscous Fluid. By **J. Prescott** 615
Effect of Surface Changes on the Photoelectric Emission of Silver and Gold. By **T. E. Clarke** 624
Some Properties of Porous Building Materials. Part III-Theory of the Absorption of Water by Porous Bodies. By **E. Madgwick** 632
Some Properties of Porous Building Materials. Part IV-The Determination of the Absorption Constants of a Homogeneous Specimen. By **E. Madgwick** 641
On the Effective Capacity of the Quadrant Electrometer. By **J. J. McHenry** 650
Continuous Atomic Matrix. By **W. R. Morgans** 664
Current Flow in a Circular Disk. By **J. H. Awberry** 674
Langmuir Probe Measurements in the Normal Copper Arc. By **Ervin H. Bramhall** 682
Special Examples of the Elastic Ring Acted upon by Equal and Equiangular Radial Forces. By **C. E. Larard** 705
On the Electrical Oscillations of a Very Short Wave-length. By **Antonio Rostagni** 733
On the Electrical Oscillations of a Very Short Wave-length. By **E. W. B. Gill** 734

NATURE

VOL. 129, NO. 3253 MARCH 5, 1932

Brief Reviews:
Origin of the Solar System. By **Herbert Dingle** 333

Letters to the Editor:
Metastable Atoms in Mercury Fluorescence. By **Lord Rayleigh** 344
Number of Free Protons in the Nucleus. By **A. J. Rutgers** 361
Cause of Twinning of Crystals. By **W. G. Burgers** 363
Polarization of Downcoming Wireless Waves. By **J. A. Radcliffe and F. W. G. White** 364
Molecular Structure of Cellulose and of Amylose. By **W. N. Haworth** 365

ZEITSCHRIFT FÜR PHYSIK

VOL. 74, NOS. 9 AND 10 MARCH 3, 1932

Effective Radius for Ionization and Electron Transfer of Argon with Respect to $Argon^{+}$. By **Franz Wolf** 575
Chemical Nature of Oxide Layers Formed by Anode Polarization on Aluminum, Zirconium, Titanium and Tantalum. By **W. G. Burgers, A. Claassen and J. Zernike** 593

Secondary Phenomena Following the Photoelectric Effect of Caesium Atoms Adsorbed on Salt Layers. By **J. H. de Boer and M. C. Teves** 604
Stationary Velocity Distribution of Electrons Diffusing in an Electric Field. By **M. Didlaukis** 624
Hyperfine Structures and Nuclear Moments of Mercury. II. By **H. Schüler and E. G. Jones** 631
Development of Internal Strains on Stretching. By **V. Caglioti and G. Sachs** 647
Law of Change of the Hall Effect in a Magnetic Field. By **C. Bellia** 655
Infrared Photography. By **Hermann Willenberg** 663
Cold Thermal Illumination. By **A. Güntherschulze and Hans Betz** 681
Electrons, Protons and the So-called Electron Magnetism. By **A. Güntherschulze** 692
Attempt to Produce Single Crystals of Binary Intermetallic Compounds. By **C. Rossi** 707
Fine Structure of the Line 4686 of He^+ in Parallel Electric and Magnetic Fields. By **D. P. Ray-Chaudhuri** 711
Remark on the Contribution of F. Seidl: The Influence of Roentgen and γ-Radiation on the Conductivity of Paraffin. By **Georges Guében** 714

DIE NATURWISSENSCHAFTEN

VOL. 20, NO. 11 — MARCH 11, 1932

Short Original Contributions
Behavior of Electron Tubes at Very High Frequencies. By **H. E. Hollmann** 181
Ratio of the Masses of the Proton and the Electron. By **C. H. Douglas Clark** 182

PHYSIKALISCHE ZEITSCHRIFT

VOL. 33, NO. 5 — MARCH 1, 1932

Complete Reviews
Photophoresis, Electro and Magneto Photophoresis. By **M. Reiss** 185–202
Papers at the 14th Meeting of the Thüringen-Sachsen-Schlesien Division of the German Physical Society
Sound Absorption Band of CO_2. By **E. Grossmann** 202
Photoelectric Coagulation of Sodium in Rock Salt. By **E. Rexer** 202–204
Photochemistry of Crystal Imperfections. By **A. Smekal** 204–206
New Method for the Measurement of After Glow in Gas Discharges. By **L. Rohde and K. Schnetzler** 206–207
Forward Motion of Electrons Due to Light. By **H. Dember** 207–208
Photoelectric Effect in Semi-Conductors. By **L. Bergmann** 209–213
On the Question of the Temperature Dependence of Predissociation. By **G. Joos and A. Herrmann** 213–214
Original Contributions
On the Measurement of Cathode Fall in Mercury Vapor Discharges. By **J. Kömmnick and Lübcke** 215–218
Frequency Dependence of the Dielectric Constant of Dilute Strong Electrolytes. By **H. Falkenhagen and E. L. Vernon** 218–221
Powerful Apparatus for the Photography of Fluorescence and Raman Spectra. By **F. Almasy** 221–223
High Frequency Conductivity of Electrolytes Measured Calorometrically. By **A. Deubner** 223–227
Remarks on the Right and Left-Hand Rules of Electrodynamics. By **W. Kaufmann** 227

ANNALEN DER PHYSIK

VOL. 12, NO. 8 — MARCH 5, 1932

Unsymmetrical Heavy Top. By **Hans Gebelein** 889
Absorption of Gases in Electrolytic Iron Melted in a Vacuum. By **F. Durau and C. H. Teckentrup** 927
Flow of Gases at Very Low Pressures through Tubes of Any Length. By **P. Clausing** 961
Querdruckelectrometer. By **H. Hermann** 990

Brownian Movement as a Factor Determining the Limit of Measurement. By **M. Czerny**. 993
Energy which Enters the Second Medium in Total Reflection. By **A. Rostagni**..... 1011
Correction to the Paper: The Dispersion of High Frequency Sound Waves in CO_2. By **H. O. Kneser**..... 1015

ANNALEN DER PHYSIK

VOL. 13, NO. 1 MARCH 11, 1932

Electron Diffraction by Pure and Passive Iron. By **W. Boas and E. Rupp**..... 1
Stationary Vortices in Shearing Fluid Motion. By **W. Lange**..... 9
Scattering of Fast Cathode Particles by Single Crystals. By **Fritz Kirchner**..... 38
Calculation of the Atom Constants e, h, N from the Rydberg Number, the Specific Charge of the Electron and the Measurement of the Short Wave-length Limit of X-ray Spectra. By **Fritz Kirchner**..... 59
Specific Volume of Water Between 20 and 650°, of Ethyl Ether and of Ethyl Alcohol Between 20 and 400° for Pressures from 1 to 2500 kg/cm². By **G. Tammann and Ad. Rühenbeck**..... 63
Magnetic Saturation. By **Ernst Czerlinsky**..... 80
Significance of Half Integral Interference in Electron Diffraction. By **E. Rupp**..... 101
Theory of Hearing in the Case of Sensitiveness to Sound through Bone Conduction. By **Georg V. Békésy**..... 111

PHYSICA

VOL. 12, NO. 3 MARCH, 1932

New Measurements on the Breakdown Potentials of Inert Gases. By **F. M. Penning**..... 65
Report Concerning Ultraviolet Irradiation..... 83

PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES

VOL. 18, NO. 3 MARCH, 1932

On the Relation Between the Expansion and the Mean Density of the Universe. By **A. Einstein and W. de Sitter**..... 213
Note on the Quantum Dynamical Correction of the Equation of State. By **Henry Margenau**..... 230
Effect of Nuclear Spin on the Radiation Excited by Electron Impact. By **W. G. Penney**.. 231
Entropy, Reversible Processes and Thermocouples. By **E. H. Kennard**..... 237
Comments on the Note of E. H. Kennard on "Entropy, Reversible Processes and Thermo-Couples." By **P. W. Bridgman**..... 242
Dynamical Systems by Continuous Spectra. By **B. O. Koopman and J. v. Neumann**..... 255
Physical Applications of the Ergodic Hypothesis. By **J. v. Neumann**..... 263
Complex Geometry and Relativity; Theory of the "Rac" Curvature. By **Edward Kasner**. 267

PHYSICS

VOL. 2, NO. 3 MARCH, 1932

Some Aspects of Electrical Prospecting Applied in Locating Oil Structures. By **Leo J. Peters and John Bardeen**..... 103
New Instrument for Measuring Very Small Differences in Gravity. By **Kenneth Hartley**.. 123
Charts for Torsion Balance Readings. By **M. M. Slotnick**..... 131
Effects of Heat Treatment on Fine Metallic Suspensions. By **N. N. Zirbel**..... 134
On the Correlation of Isogeothermal Surfaces with the Rock Strata. By **C. E. Van Orstrand**..... 139
Geothermal Gradient Determinations in the Lake Superior Copper Mines. By **L. R. Ingersoll**..... 154
Velocity of Elastic Waves in Granite. By **L. Don Leet and W. Maurice Ewing**..... 160
Asymmetry of Sound Velocity in Stratified Formations. By **Burton McCollum and F. A. Snell**..... 174
Calculation of the Motion of the Ground from Seismograms. By **H. A. Wilson**..... 186